

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2003, 11:55:48 ; Search time 41 Seconds  
(without alignments)  
58.071 Million cell updates/sec

Title: US-09-807-610-11

Perfect score: 80

Sequence: 1 ALADLYEEGGGGGGE 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq 19Jun03.\*

```

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	80	100.0	15	21	AA98151
2	80	100.0	24	19	AAW65063
3	80	100.0	180	17	AA91361
4	80	100.0	180	20	AA933278
5	80	100.0	180	23	ABP52019
6	80	100.0	180	23	ABP52033
7	80	100.0	180	24	AAE33566
8	72	90.0	19	19	AAW65061
9	72	90.0	24	19	AAW65064

10	67	83.8	21	17	AA91360	Intracellular IL-1
11	65	81.2	19	19	AAW65060	ICE binding peptid
12	64	80.0	19	19	AAW65062	ICE binding peptid
13	53	66.2	24	15	AA960407	Antiproliferative
14	51	63.7	48	22	ABB29117	Peptide #1768 enco
15	51	63.7	48	22	AAW55071	Human brain expres
16	51	63.7	48	22	AAW67464	Human bone marrow
17	51	63.7	48	22	AAW15286	Peptide #1720 enco
18	51	63.7	77	22	ABG49761	Human liver peptid
19	51	63.7	77	22	ABB29752	Peptide #2403 enco
20	51	63.7	77	22	ABB34925	Peptide #2431 enco
21	51	63.7	77	22	ABB20336	Protein #2335 enco
22	51	63.7	77	22	AAW55733	Human brain expres
23	51	63.7	77	22	AAW68114	Human bone marrow
24	51	63.7	77	22	AAW15935	Peptide #2369 enco
25	51	63.7	77	22	AAW28438	Peptide #2475 enco
26	51	63.7	77	22	AAW03671	Peptide #2353 enco
27	51	63.7	77	23	ABG37637	Human peptide enco
28	51	63.7	258	22	AAE00799	Human Her-2/neu ov
29	51	63.7	260	22	ABW12434	Human bone marrow
30	51	63.7	264	22	ABG05522	Novel human diagno
31	51	63.7	434	22	ABG03902	Novel human diagno
32	51	63.7	898	22	ABG13068	Novel human diagno
33	51	63.7	1118	22	ABG16323	Novel human diagno
34	50	62.5	977	23	AAE22220	Human toll like re
35	48	60.0	75	21	AAE23802	Arabidopsis thalia
36	48	60.0	83	21	AAE23801	Arabidopsis thalia
37	48	60.0	148	21	AAE23800	Arabidopsis thalia
38	48	60.0	219	22	ABG11553	Novel human diagno
39	48	60.0	290	22	AAO05792	Human polypeptide
40	48	60.0	325	23	AAU93031	Arabidopsis transc
41	48	60.0	940	21	AAW58041	Escherichia coli U
42	48	60.0	940	22	AAU34854	E. coli cellular p
43	48	60.0	940	22	AAU69758	Escherichia coli U
44	48	60.0	941	22	AAU38225	Salmonella typhi c
45	47	58.8	261	22	AAU02927	Angiotensin conver

#### ALIGNMENTS

##### RESULT 1

AA98151  
ID AA98151 standard; Protein; 15 AA.

XX AA98151;

AC AA98151;

XX 22-AUG-2000 (first entry)

DT Human icIL-1raII amino acid fragment.

DE Expression vector; icIL-1raII; Interleukin-1; IL-1; growth hormone;

DE signal peptide; intracellular IL-1 receptor antagonist type II; cytokine;

DE human; IL-1 overexpression.

XX Homo sapiens.

OS Homo sapiens.

XX WO200022146-A1.

PN 20-APR-2000.

XX 14-OCT-1999; 99WO-IL00543.

XX 14-OCT-1998; 98IL-0126562.

XX (INTE-) INTERPHARM LAB LTD.

XX (ISTF) ARS APPLIED RES SYSTEMS HOLDING NV.

XX Amitai H, Chitlaru E;

XX WPI; 2000-317997/27.

XX A new expression vector for intracellular IL-1 receptor antagonist type

PT

II joined to a signal peptide is useful for reducing IL-1 in a patient and allows secretion of the intracellular molecule

Example 11; Page 17; 35pp; English.

This invention relates to an expression vector comprising DNA encoding a signal peptide of a protein normally expressed and secreted by human cells (e.g. human growth hormone signal peptide) joined to DNA encoding intracellular interleukin 1 (IL-1) receptor antagonist type II (icIL-1raII) and operably linked to a promoter sequence. The expression results in a signal peptide fused in frame to icIL-1raII. IL-1ra is a unique IL-1 receptor antagonist, that inhibits signalling from the functional IL-1 receptor. icIL-1raII inhibits exogenous IL-1 dependent responses, and is termed intracellular due to its lack of signal peptide. The expression vector can be used in a process to engineer proteins to be secreted. The expression vector and resulting recombinant proteins to be secreted are used to reduce IL-1 in a patient, either by direct administration of the glycosylated icIL-1raII, or by gene therapy using the vector. The present sequence represents a fragment of icIL-1raII, obtained in protein sequence analysis to prove that the expression vector can be used to create secreted mature icIL-1ra type II protein.

Sequence -15 AA;

Query Match 100.0%; Score 80; DB 21; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 1 ALADLYEKGGGGGG 15  
 |||||  
 Db 1 ALADLYEKGGGGGG 15

RESULT 2

AAW65063

ID AAW65063 standard; peptide; 24 AA.

AC AAW65063;

XX 15-SBP-1998 (first entry)

DE ICE binding peptide C.

XX Interleukin-1-beta converting enzyme; ICE; autoimmune disease; infection;  
 inflammatory diseases; rheumatoid arthritis; septic shock; leukaemia;  
 multiple sclerosis; diagnosis; cysteine protease; inhibitor; cytokine.  
 Synthetic.  
 Homo sapiens.  
 WO9818823-A1.  
 07-MAY-1998.

XX 31-OCT-1996; 96WO-EP04738.

XX 31-OCT-1996; 96WO-EP04738.

PA (ISTP) ARS APPLIED RES SYSTEMS HOLDING NV.

PI Introna M, Mantovani A, Muzio M;  
 WPI; 1998-272138/24.

XX Peptide binding and inhibiting interleukin-1 $\beta$ -converting enzyme,  
 ICE - useful in treating pathologies requiring inhibition of ICE or  
 ICE family enzymes, e.g. autoimmune or inflammatory diseases etc.  
 Claim 4; Figure 3; 32pp; English.

XX AAW65060-W65067 are peptides capable of binding to interleukin-1-beta  
 converting enzyme (ICE) and/or enzymes of the ICE family. Such peptides  
 can be used medically, especially to prepare pharmaceutical compositions

CC which can be administered prophylactically or therapeutically to  
 CC patients with pathologies requiring ICE inhibition and/or inhibition of  
 CC enzymes of the ICE family, e.g. autoimmune diseases, lethal bacterial or  
 CC viral infections or inflammatory diseases. Specific examples of such  
 CC pathologies include rheumatoid arthritis, septic shock, leukaemia,  
 CC multiple sclerosis etc. The peptide may also be used in the diagnosis of  
 CC such pathologies. ICE is a heterodimeric cysteine protease which cleaves  
 CC interleukin-1 beta precursor protein (pIL-1 $\beta$ ) during interleukin-1  
 CC beta (IL-1 $\beta$ ) synthesis. Inhibition of ICE can thus inhibit synthesis  
 CC of IL-1 $\beta$ . Interleukin-1's are potent inflammatory and pyrogenic  
 CC cytokines that normally have beneficial effects but also act in many  
 CC human pathologies, especially on the immune response and on inflammatory  
 CC processes.

SQ Sequence 24 AA;

Query Match 100.0%; Score 80; DB 19; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ALADLYEKGGGGGG 15  
 |||||  
 Db 2 ALADLYEKGGGGGG 16

RESULT 3

AAW91361

ID AAW91361 standard; Protein; 180 AA.

AC AAW91361;

XX 25-MAR-2003 (updated)

DT 22-OCT-1996 (first entry)

XX Intracellular IL-1 receptor antagonist type II.

XX Intracellular IL-1 receptor antagonist; icIL-1ra;  
 secreted IL-1 receptor antagonist; sIL-1ra;  
 interleukin; IL-1a; IL-1b; auto-immune disease.  
 Homo sapiens.  
 WO9612022-A1.  
 25-APR-1996.

XX 12-OCT-1995; 95WO-EP04023.

XX 13-OCT-1994; 94IT-MI02097.

PA (ISTP) ARS APPLIED RES SYST HOLDING NV.

PI Colotta F, Mantovani A, Muzio M;  
 WPI; 1996-222008/22.  
 DR N-PSDB; AAT15099.

XX IL-1 receptor antagonist active against IL-1a and -1b - for  
 PT treating, preventing or diagnosing auto-immune diseases  
 PS Claim 2; Page 22-23; 36pp; English.

XX An new IL-1 receptor antagonist includes the sequence given in AAW91360.  
 CC The complete icIL-1raII is given in AAT15099. The protein is  
 CC expressed by DNA similar to that encoding the known receptor  
 CC inhibitor icIL-1ra, but having a 63 bp insert between the first  
 CC icIL-1ra specific exon and the internal acceptor site of the first  
 CC exon of sIL-1ra.  
 CC (Updated on 25-MAR-2003 to correct PR field.)

XX Sequence 180 AA;

Query Match 100.0%; Score 80; DB 17; Length 180;

Best Local Similarity 100.0%; Pred. No. 0.00023; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ALADLYEEGGGGGGE 15  
Db 2 ALADLYEEGGGGGGE 16

RESULT 4  
AAY33278  
ID AAY33278 standard; protein; 180 AA.

AC AAY33278;

DT 23-NOV-1999 (first entry)

DE ICIRAPII conserved peptide fragment.

KW Interleukin-1; IL-1; cytokine; inhibitor; antagonist; receptor; trauma;  
KW anti-inflammatory; anti-infectious; apoptotic; inflammatory; infection;  
KW catabolic; degeneration; chronic inflammation; autoimmune disease; IL-1b;  
KW nervous system; motor system; inhibitor; erythropoiesis; chondrocyte;  
KW rheumatic patient; collagen synthesis; hepatocyte; cell-cell adhesion;  
KW cartilage degradation; nerve regeneration; necrosis induction; apoptosis;  
KW intracellular IL-1 receptor agonist; ICIRAPII.

XX Unidentified.

OS OS

PN DE19753753-A1.

PD 26-AUG-1999.

PF 04-DEC-1997; 97DE-1053753.

PR 04-DEC-1997; 97DE-1053753.

PA (ORTH-) ORTHOGEN GENTECNOLOGIE GMBH.

PI Meijer H, Wehling P, Reinecke J;

XX WPI; 1999-479917/41.

DR New recombinant nucleic acid encoding proteins for modifying effects of

PT cytokine systems, e.g. for treating infection or trauma

XX Disclosure; Fig 4; 8pp; German.

CC This invention describes novel recombinant nucleic acids (I) that encode  
CC proteins (II) which modify the effects of cytokines or their  
CC inhibitors/antagonists or their receptors. The products of the invention  
CC have anti-inflammatory, anti-infectious and apoptotic activity. (I) and  
CC (II) are particularly used to modulate activities of the human  
CC interleukin-1 (IL-1) system which is involved in many inflammatory and  
CC catabolic processes, e.g. in infection, trauma, degeneration and chronic  
CC inflammation (autoimmune diseases) of the nervous and motor systems and  
CC internal organs. In particular, (II) are used in cases where IL-1  
CC inhibits erythropoiesis in rheumatic patients, inhibits collagen  
CC synthesis by chondrocytes, inhibits growth of hepatocytes, promotes  
CC cell-cell adhesion, immune reactions, degradation of cartilage, nerve  
CC regeneration, and induction of necrosis or apoptosis in the central  
CC nervous system. This sequence represents a fragment of an intracellular  
CC IL-1 receptor agonist (ICIRAPII) which is used to describe the method of  
CC the invention.

XX Sequence 180 AA;

Query Match 100.0%; Score 80; DB 20; Length 180;

Best Local Similarity 100.0%; Pred. No. 0.00023; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ALADLYEEGGGGGGE 15  
Db 2 ALADLYEEGGGGGGE 16

RESULT 5  
ABP52019  
ID ABP52019 standard; Protein; 180 AA.

XX AC ABP52019;

XX DT 10-OCT-2002 (first entry)

XX NOVINTRA A homologous amino acid sequence SEQ ID NO:62.

KW Human; transmembrane protein; neuromedin protein; gonadotropin protein;  
KW interleukin-1 receptor antagonist; interleukin-1 epsilon; NOVI;  
KW IL-1 epsilon; IL-1 receptor antagonist; lung disease; nootropic;  
KW cytosolic; neuroprotective; antiinflammatory; antibacterial;  
KW immunosuppressive; cerebroprotective; antidiabetic; antiarthritic;  
KW antiasthmatic; antiallergic; gene therapy; antibody-based therapy;  
KW cell signalling disorder; haematopoietic disorder; endocrine; muscle;  
KW neurodegenerative disorder; neurological disorder; cancer; melanoma;  
KW central nervous system cancer; reproductive development disorder; asthma;  
KW metabolic function disorder; bone metabolism; structure disorder; stroke;  
KW inflammatory response disorder; immune regulation disorder; septic shock;  
KW diabetes; arthritis; lung cancer; emphysema; allergic lung irritation;  
KW lung inflammation.

XX Homo sapiens.

XX US2002068279-A1.

XX 06-JUN-2002.

XX 05-DEC-2000; 2000US-0730617.

XX 06-DEC-1999; 99US-169056P.

XX 09-DEC-1999; 99US-169866P.

XX 09-DEC-1999; 99US-169886P.

XX 10-DEC-1999; 99US-170252P.

XX 12-JAN-2000; 2000US-175740P.

XX (CURA-) CURAGEN CORP.

XX Burgess C, Prayaga SK, Shinkets RA, Rastelli L, Zerhusen B;

XX Mezes P;

XX WPI; 2002-582472/62.

XX New NOVX proteins for diagnosing or treating cell signaling, immune

XX response, hematopoietic, neurodegenerative, muscle, endocrine, bone,

XX and reproductive development disorders

XX Disclosure; Fig 11; 110pp; English.

CC The present invention describes an isolated NOVX polypeptide, chosen from  
CC human transmembrane (NOVINTRA), neuromedin (NOVNEUR), gonadotropin  
CC (NOVGON), interleukin-1 (IL-1) receptor antagonist (NOVINTRA A and B),  
CC and IL-1 epsilon proteins. NOVX polypeptides have nootropic, cytosolic,  
CC neuroprotective, antiinflammatory, antibacterial, immunosuppressive,  
CC cerebroprotective, antidiabetic, antiarthritic, antiasthmatic and  
CC antiallergic activities, and can be used in gene therapy and antibody-  
CC based therapy. NOVX polypeptides, nucleic acid (I) encoding them and an  
CC antibody (III) that binds the polypeptide, are useful for treating or  
CC preventing a NOVX protein-associated disorder in humans. NOVNEUR can be  
CC used in the treatment of a cell signalling disorder, such as, a  
CC haematopoietic disorder or a neurodegenerative disorder. NOVNEUR can be  
CC used in the treatment of an endocrine, muscle, neurological disorder,  
CC central nervous system cancer, breast, colon, ovarian, kidney, prostate  
CC or thyroid cancer. NOVGON can be used in the treatment of a reproductive  
CC development disorder, metabolic function disorder or melanoma. NOVINTRA  
CC proteins can be used in the treatment of and a bone metabolism or  
CC structure disorder, an inflammatory response disorder, an immune  
CC regulation disorder, septic shock, stroke, diabetes, arthritis or  
CC cancer. An agent which modulates the expression or activity of a human

CC IL-1 epsilon protein is useful for treating a lung disease such as lung  
 CC cancer, asthma, emphysema, allergic lung irritation and lung inflammation  
 CC in a mammal. ABQ73996 to ABQ74027 and ABP51981 to ABP52048 represent  
 CC sequences used in the exemplification of the present invention.

XX Sequence 180 AA;

Query Match 100.0%; Score 80; DB 23; Length 180;  
 Best Local Similarity 100.0%; Pred. No. 0.00023;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ALADLYEKGGGGGG 15  
 |||||  
 Db 2 ALADLYEKGGGGGG 16

# RESULT 6

ID ABP52033 standard; Protein; 180 AA.

XX AC ABP52033;

DT 10-OCT-2002 (first entry)

DE NOVINTRA B homologous amino acid sequence SEQ ID NO:79.

XX Human; transmembrane protein; neuromedin protein; gonadotropin protein;  
 KW interleukin-1 receptor antagonist; interleukin-1 epsilon; NOVX;  
 KW IL-1 epsilon; IL-1 receptor antagonist; lung disease; neutrotic;  
 KW cytostatic; neuroprotective; antiinflammatory; antibacterial;  
 KW immunosuppressive; cerebroprotective; antidiabetic; antiarthritic;  
 KW antiasthmatic; antiallergic; gene therapy; antibody-based therapy;  
 KW cell signalling disorder; haematopoietic disorder; endocrine; muscle;  
 KW neurodegenerative disorder; neurological disorder; cancer; melanoma;  
 KW central nervous system cancer; reproductive development disorder; asthma;  
 KW metabolic function disorder; bone metabolism; structure disorder; stroke;  
 KW inflammatory response disorder; immune regulation disorder; septic shock;  
 KW diabetes; arthritis; lung cancer; emphysema; allergic lung irritation;  
 KW lung inflammation.

XX Homo sapiens.

OS US2002068279-A1.

FN 06-JUN-2002.

XX 05-DEC-2000; 2000US-0730617.

XX 06-DEC-1999; 99US-169056P.

PR 09-DEC-1999; 99US-169866P.

PR 09-DEC-1999; 99US-169886P.

PR 10-DEC-1999; 99US-170252P.

PR 12-JAN-2000; 2000US-175740P.

XX (CURA-) CURAGEN CORP.

XX Burgess C, Prayaga SK, Shimkets RA, Rastelli L, Zerhusen B;

PI Mezes P;

XX WPI; 2002-582472/62.

XX New NOVX proteins for diagnosing or treating cell signaling, immune  
 PT response, hematopoietic, neurodegenerative, muscle, endocrine, bone,  
 PT and reproductive development disorders -

XX Disclosure; Fig 14; 110pp; English.

XX The present invention describes an isolated NOVX polypeptide, chosen from  
 CC human transmembrane (NOVTRAM), neuromedin (NOVNEUR), gonadotropin  
 CC (NOVCON), interleukin-1 (IL-1) receptor antagonist (NOVINTRA A and B),  
 CC and IL-1 epsilon proteins. NOVX polypeptides have neutrotic, cycostatic,  
 CC neuroprotective, antiinflammatory, antibacterial, immunosuppressive,  
 CC cerebroprotective, antidiabetic, antiarthritic, antiasthmatic and

CC antiallergic activities, and can be used in gene therapy and antibody-  
 CC based therapy. NOVX polypeptides, nucleic acid (I) encoding them and an  
 CC antibody (III) that binds the polypeptide, are useful for treating or  
 CC preventing a NOVX protein-associated disorder in humans. NOVTRAM can be  
 CC used in the treatment of a cell signalling disorder, such as, a  
 CC haematopoietic disorder or a neurodegenerative disorder. NOVNEUR can be  
 CC used in the treatment of an endocrine, muscle, neurologic disorder,  
 CC central nervous system cancer, breast, colon, ovarian, kidney, prostate  
 CC or thyroid cancer. NOVCON can be used in the treatment of a reproductive  
 CC development disorder, metabolic function disorder or melanoma. NOVINTRA  
 CC proteins can be used in the treatment of a bone metabolism or  
 CC structure disorder, an inflammatory response disorder, an immune  
 CC regulation disorder, septic shock, stroke, diabetes, arthritis or  
 CC cancer. An agent which modulates the expression or activity of a human  
 CC IL-1 epsilon protein is useful for treating a lung disease such as lung  
 CC cancer, asthma, emphysema, allergic lung irritation and lung inflammation  
 CC in a mammal. ABQ73996 to ABQ74027 and ABP51981 to ABP52048 represent  
 CC sequences used in the exemplification of the present invention.

XX Sequence 180 AA;

Query Match 100.0%; Score 80; DB 23; Length 180;  
 Best Local Similarity 100.0%; Pred. No. 0.00023;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ALADLYEKGGGGGG 15  
 |||||  
 Db 2 ALADLYEKGGGGGG 16

# RESULT 7

AAE33566

ID AAE33566 standard; Protein; 180 AA.

AC AAE33566;

DT 16-APR-2003 (first entry)

DE Intracellular interleukin-1RN (IL-1RN).

XX Drug screening; fungicide; gene therapy; antibacterial; infection;  
 KW virucide; interleukin-1RN; IL-1RN.

XX Unidentified.

XX WO2002101015-A2.

XX 19-DEC-2002.

XX 11-JUN-2002; 2002WO-US18346.

XX 11-JUN-2001; 2001US-297305P.

XX (INTE-) INTERLEUKIN GENETICS INC.

XX Dower S, Duff GW;

XX WPI; 2003-148793/14.

XX New detection reagent, useful for monitoring molecular assembly events  
 PT to permit the dissection of genetic and non-genetic influences on  
 PT biological activity, comprises an interactive sensor pair -

XX Claim 17; Fig 8; 56pp; English.

XX The invention relates to methods, compositions and apparatus for  
 CC monitoring molecular assembly events. It also relates to a detection  
 CC reagent comprising an interactive sensor pair. The detection reagent is  
 CC useful for monitoring molecular assembly events to permit the dissection  
 CC of genetic and non-genetic influences on a particular biological  
 CC activity. The method is useful for linking genetic variations to  
 CC molecular and physiological events, drug screening, diagnostics, therapy  
 CC selection and dosing, patient monitoring or environmental safety. The

CC interactive sensor pairs may be used to screen for and identify novel  
CC agonists and antagonists or other molecules that modulate a biological  
CC activity. The method is also useful for selecting an appropriate  
CC targeted therapeutic for a subject having an infection, including  
CC viral, bacterial or fungal infection. It is also used in gene therapy.  
CC The present sequence is an intracellular interleukin-1RN (IL-1RN). This  
CC sequence is used to illustrate the method of the invention.

XX SQ Sequence 180 AA;

Query Match 100.0%; Score 80; DB 24; Length 180;

Best Local Similarity 100.0%; Pred. No. 0.00023;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ALADLYERGGGGGE 15

DB 2 ALADLYERGGGGGE 16

RESULT 8

AAW65061

ID AAW65061 standard; peptide; 19 AA.

XX AC AAW65061;

DT 15-SEP-1998 (first entry)

DE ICE binding peptide #2.

XX Interleukin-1-beta converting enzyme; ICE; autoimmune disease; infection;  
KW inflammatory diseases; rheumatoid arthritis; septic shock; leukaemia;  
KW multiple sclerosis; diagnosis; cysteine protease; inhibitor; cytokine.  
XX Synthetic.  
OS Homo sapiens.

OS WO9818823-A1.

PN 07-MAY-1998.

PD 31-OCT-1996; 96WO-EP04738.

PP 31-OCT-1996; 96WO-EP04738.

PR (ISTP) ARS APPLIED RES SYSTEMS HOLDING NV.

PA Introna M, Mantovani A, Muzio M;

PI WPI; 1998-272138/24.

DR Peptide binding and inhibiting interleukin-1'b-converting enzyme,

PT ICE - useful in treating pathologies requiring inhibition of ICE or

PT ICE family enzymes, e.g. autoimmune or inflammatory diseases etc.

XX Claim 2; Page 14; 32pp; English.

XX AAW65060-W65067 are peptides capable of binding to interleukin-1-beta  
CC converting enzyme (ICE) and/or enzymes of the ICE family. Such peptides  
CC can be used medicinally, especially to prepare pharmaceutical compositions  
CC which can be administered prophylactically or therapeutically to  
CC patients with pathologies requiring ICE inhibition and/or inhibition of  
CC enzymes of the ICE family, e.g. autoimmune diseases, lethal bacterial or  
CC viral infections or inflammatory diseases. Specific examples of such  
CC pathologies include rheumatoid arthritis, septic shock, leukaemia,  
CC multiple sclerosis etc. The peptide may also be used in the diagnosis of  
CC interleukin-1 beta precursor protein (pIL-1-beta) during interleukin-1  
CC beta (IL-1-beta) synthesis. Inhibition of ICE can thus inhibit synthesis  
CC of IL-1-beta. Interleukin-1's are potent inflammatory and pyrogenic  
CC cytokines that normally have beneficial effects but also act in many  
CC human pathologies, especially on the immune response and on inflammatory  
CC processes.

SQ Sequence 19 AA;

Query Match 90.0%; Score 72; DB 19; Length 19;  
Best Local Similarity 100.0%; Pred. No. 0.00032;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ADLYERGGGGGE 15

DB 1 ADLYERGGGGGE 13

RESULT 9

AAW65064

ID AAW65064 standard; peptide; 24 AA.

XX AC AAW65064;

DT 15-SEP-1998 (first entry)

DE ICE binding peptide S.

XX Interleukin-1-beta converting enzyme; ICE; autoimmune disease; infection;  
KW inflammatory diseases; rheumatoid arthritis; septic shock; leukaemia;  
KW multiple sclerosis; diagnosis; cysteine protease; inhibitor; cytokine.  
XX Synthetic.  
OS Homo sapiens.

OS WO9818823-A1.

PN 07-MAY-1998.

PD 31-OCT-1996; 96WO-EP04738.

PP 31-OCT-1996; 96WO-EP04738.

PR (ISTP) ARS APPLIED RES SYSTEMS HOLDING NV.

PA Introna M, Mantovani A, Muzio M;

PI WPI; 1998-272138/24.

DR Peptide binding and inhibiting interleukin-1'b-converting enzyme,  
PT ICE - useful in treating pathologies requiring inhibition of ICE or  
PT ICE family enzymes, e.g. autoimmune or inflammatory diseases etc.  
XX Claim 5; Figure 3; 32pp; English.

XX AAW65060-W65067 are peptides capable of binding to interleukin-1-beta  
CC converting enzyme (ICE) and/or enzymes of the ICE family. Such peptides  
CC can be used medicinally, especially to prepare pharmaceutical compositions  
CC which can be administered prophylactically or therapeutically to  
CC patients with pathologies requiring ICE inhibition and/or inhibition of  
CC enzymes of the ICE family, e.g. autoimmune diseases, lethal bacterial or  
CC viral infections or inflammatory diseases. Specific examples of such  
CC pathologies include rheumatoid arthritis, septic shock, leukaemia,  
CC multiple sclerosis etc. The peptide may also be used in the diagnosis of  
CC interleukin-1 beta precursor protein (pIL-1-beta) during interleukin-1  
CC beta (IL-1-beta) synthesis. Inhibition of ICE can thus inhibit synthesis  
CC of IL-1-beta. Interleukin-1's are potent inflammatory and pyrogenic  
CC cytokines that normally have beneficial effects but also act in many  
CC human pathologies, especially on the immune response and on inflammatory  
CC processes.

SQ Sequence 24 AA;

Query Match 90.0%; Score 72; DB 19; Length 24;  
Best Local Similarity 93.3%; Pred. No. 0.00042;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ALADLYERGGGGGE 15

DB 1 ALADLYERGGGGGE 13

```

Db      2 ALAALYBEGGGGGG 16

RESULT 10
AAR91360
ID AAR91360 standard; Peptide; 21 AA.
XX AC AAR91360;
XX AC
XX DT 25-MAR-2003 (updated)
XX DT 22-OCT-1996 (first entry)
XX DE
XX DE Intracellular IL-1 receptor antagonist type II fragment.
XX KW Intracellular IL-1 receptor antagonist; icIL-1ra;
XX KW secreted IL-1 receptor antagonist; sIL-1ra;
XX KW interleukin; IL-1a; IL-1b; auto-immune disease.
XX OS Homo sapiens.
XX PN WO9612022-A1.
XX PD 25-APR-1996.
XX PP 12-OCT-1995; 95WO-EP04023.
XX PR 13-OCT-1994; 94IT-MI02097.
XX PA (ISTF) ARS APPLIED RES SYST HOLDING NV.
XX PI Colotta F, Mantovani A, Muzio M;
XX OR WPI, 1996-222008/22.
XX PT IL-1 receptor antagonist active against IL-1a and -1b - for
XX PT treating, preventing or diagnosing auto-immune diseases
XX PS Claim 1; Page 21; 36pp; English.
XX CC An new IL-1 receptor antagonist includes the sequence given in AAR91360.
XX CC The complete icIL-1ra1 is given in AAT15099. The protein is
XX CC expressed by DNA similar to that encoding the known receptor
XX CC inhibitor icIL-1ra, but having a 63 bp insert between the first
XX CC icIL-1ra specific exon and the internal acceptor site of the first
XX CC exon of sIL-1ra
XX CC (Updated on 25-MAR-2003 to correct PR field.)
XX SQ Sequence 21 AA;
Query Match 83.8%; Score 67; DB 17; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DY 3 ADLYBEGGGGGG 14
| | | | |
DB 1 ADLYBEGGGGGG 12

RESULT 11
AAW65060
ID AAW65060 standard; peptide; 19 AA.
XX AC AAW65060;
XX DT 15-SEP-1998 (first entry)
XX DE ICB binding peptide #1.
XX KW Interleukin-1-beta converting enzyme; ICB; autoimmune disease; infection;
XX KW inflammatory disease; rheumatoid arthritis; septic shock; leukaemia;
XX KW multiple sclerosis; diagnosis; cysteine protease; inhibitor; cytokine.
XX OS Synthetic.
XX OS Homo sapiens

OS Homo sapiens
XX FH Key Location/Qualifiers
FT Misc-difference 2 /note= "can be either Ala or more preferably Asp"
FT Misc-difference 19 /note= "can be either Ala or more preferably Asp"
XX PN WO9818823-A1.
XX PD 07-MAY-1998.
XX PP 31-OCT-1996; 96WO-EP04738.
XX PR 31-OCT-1996; 96WO-EP04738.
XX PA (ISTF) ARS APPLIED RES SYSTEMS HOLDING NV.
XX PI Introna M, Mantovani A, Muzio M;
XX OR WPI; 1998-272138/24.
XX PT Peptide binding and inhibiting interleukin-1'b-converting enzyme,
XX PT ICB - useful in treating pathologies requiring inhibition of ICB or
XX PT ICB family enzymes, e.g. autoimmune or inflammatory diseases etc.
XX PS Claim 1; Page 14; 32pp; English.
XX CC AAW65060-W65067 are peptides capable of binding to interleukin-1-beta
XX CC converting enzyme (ICB) and/or enzymes of the ICB family. Such peptides
XX CC can be used medically, especially to prepare pharmaceutical compositions
XX CC which can be administered prophylactically or therapeutically to
XX CC patients with pathologies requiring ICB inhibition and/or inhibition of
XX CC enzymes of the ICB family, e.g. autoimmune diseases, lethal bacterial or
XX CC viral infections or inflammatory diseases. Specific examples of such
XX CC pathologies include rheumatoid arthritis, septic shock, leukaemia,
XX CC multiple sclerosis etc. The peptide may also be used in the diagnosis of
XX CC such pathologies. ICB is a heterodimeric cysteine protease which cleaves
XX CC interleukin-1 beta precursor protein (pIL-1-beta) during interleukin-1
XX CC beta (IL-1-beta) synthesis. Inhibition of ICB can thus inhibit synthesis
XX CC of IL-1-beta. Interleukin-1's are potent inflammatory and pyrogenic
XX CC cytokines that normally have beneficial effects but also act in many
XX CC human pathologies, especially on the immune response and on inflammatory
XX CC processes.
XX SQ Sequence 19 AA;
Query Match 81.2%; Score 65; DB 19; Length 19;
Best Local Similarity 92.3%; Pred. No. 0.0036;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 ADLYBEGGGGGG 15
| | | | |
DB 1 ADLYBEGGGGGG 13

RESULT 12
AAW65062
ID AAW65062 standard; peptide; 19 AA.
XX AC AAW65062;
XX DT 15-SEP-1998 (first entry)
XX DE ICB binding peptide #3.
XX KW Interleukin-1-beta converting enzyme; ICB; autoimmune disease; infection;
XX KW inflammatory diseases; rheumatoid arthritis; septic shock; leukaemia;
XX KW multiple sclerosis; diagnosis; cysteine protease; inhibitor; cytokine.
XX OS Synthetic.
XX OS Homo sapiens

```

PN WO9818823-A1.  
 XX 07-MAY-1998.  
 PD 31-OCT-1996; 96WO-EP04738.  
 XX 31-OCT-1996; 96WO-EP04738.  
 PR (1STF ) ARS APPLIED RES SYSTEMS HOLDING NV.  
 XX Introna M, Mantovani A, Muzio M;  
 XX WPI; 1998-272138/24.  
 DR Peptide binding and inhibiting interleukin-1<sup>b</sup>-converting enzyme,  
 XX ICE - useful in treating pathologies requiring inhibition of ICE or  
 XX ICE family enzymes, e.g. autoimmune or inflammatory diseases etc.  
 PS Claim 3; Page 14; 32pp; English.  
 XX AAM65060-W65067 are peptides capable of binding to interleukin-1-beta  
 CC converting enzyme (ICE) and/or enzymes of the ICE family. Such peptides  
 CC can be used medicinally, especially to prepare pharmaceutical compositions  
 CC which can be administered prophylactically or therapeutically to  
 CC patients with pathologies requiring ICE inhibition and/or inhibition of  
 CC enzymes of the ICE family, e.g. autoimmune diseases, lethal bacterial or  
 CC viral infections or inflammatory diseases. Specific examples of such  
 CC pathologies include rheumatoid arthritis, septic shock, leukaemia,  
 CC multiple sclerosis etc. The peptide may also be used in the diagnosis of  
 CC such pathologies. ICE is a heterodimeric cysteine protease which cleaves  
 CC interleukin-1 beta precursor protein (pIL-1-beta) during interleukin-1  
 CC beta (IL-1-beta) synthesis. Inhibition of ICE can thus inhibit synthesis  
 CC of IL-1-beta. Interleukin-1's are potent inflammatory and pyrogenic  
 CC cytokines that normally have beneficial effects but also act in many  
 CC human pathologies, especially on the immune response and on inflammatory  
 CC processes.  
 XX Sequence 19 AA;  
 SQ

Query Match 80.0%; Score 64; DB 19; Length 19;  
 Best Local Similarity 92.3%; Pred. No. 0.0051;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ADLYREGGGGGG 15  
 DB 1 ADLYREGGGGGG 13

RESULT 13  
 AAR60407  
 ID AAR60407 standard; peptide; 24 AA.  
 XX AAR60407;  
 XX 25-MAR-2003 (updated)  
 DT 30-MAR-1995 (first entry)  
 XX Antiproliferative peptide C (repeat) to human B-cell lymphoma.  
 DE antiproliferative; transplant; B-cell lymphoma line SUP-B8; Burkitt's;  
 KW inhibit clonal expansion; induce apoptosis; anti-idiotypic; IgM lambda;  
 KW inhibit cell proliferation; peptidomimetics; cell surface receptor;  
 KW immunoglobulin superfamily; treatment; neoplasia; identification;  
 KW induce replication; therapy; clonal anergy; modulate tyrosine kinase.  
 XX Synthetic.  
 OS  
 XX Key Location/Qualifiers  
 PH Misc-difference 1...9 /note= "direct repeat"  
 FT Misc-difference 16..24 /note= "direct repeat"  
 FT  
 FT  
 FT

PN WO9418345-A1.  
 XX 18-AUG-1994.  
 PD 04-FEB-1994; 94WO-US01319.  
 XX 05-FEB-1993; 93US-0014426.  
 PR 15-NOV-1993; 93US-0153341.  
 XX (AFFY-) AFFYMAX TECHNOLOGIES NV.  
 XX (STRD ) UNIV LELAND STANFORD JUNIOR.  
 PA Bhatt RR, Dower WJ, Levy R, Renschler MP;  
 XX WPI; 1994-279762/34.  
 DR Identifying anti-proliferative peptide(s) which specifically bind  
 XX to immunoglobulin super-family species idiotype - esp. to inhibit  
 PT B-cell lymphoma and leukocytic leukaemia cell proliferation, for  
 PT anti-idiotypic therapy  
 XX Claim 7; Page 49; 69pp; English.  
 PS AAR60400-73 are peptide ligands which bind to purified IgM lambda  
 CC receptor of the human Burkitt's lymphoma cell line SUP-B8.  
 CC The peptides were identified with the use of filamentous phage  
 CC libraries displaying random peptides. Corresponding synthetic  
 CC peptides bound specifically to this Ig receptor, and blocked the  
 CC binding of an anti-idiotypic antibody. The ligands, when conjugated  
 CC to form dimers or tetramers, induced cell death by apoptosis in  
 CC vitro at nanomolar concentrations. This effect was associated with  
 CC the specific stimulation of intracellular protein tyrosine  
 CC phosphorylation. The peptides of the invention can be used individually,  
 CC as complexes of cross-linked peptides or can be conjugated to deliver  
 CC toxins or radionuclides to neoplastic cells bearing the specific Ig  
 CC receptor.  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 XX Sequence 24 AA;  
 SQ

Query Match 66.2%; Score 53; DB 15; Length 24;  
 Best Local Similarity 81.8%; Pred. No. 0.29;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 DLYREGGGGGG 14  
 DB 5 DLYREGGGGGG 15

RESULT 14  
 ABB29117  
 ID ABB29117 standard; Peptide; 48 AA.  
 XX ABB29117;  
 XX 01-FEB-2002 (first entry)  
 DT Peptide #1768 encoded by breast cell single exon nucleic acid probe.  
 DE Human; microarray; single exon probe; gene expression; breast;  
 KW disease; cancer.  
 KW Homo sapiens.  
 OS WO200157271-A2.  
 XX 09-AUG-2001.  
 PD 30-JAN-2001; 2001WO-US00662.  
 XX 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes,  
 PT useful for measuring gene expression in sample derived from human  
 PT breast, comprises number of single exon nucleic acid probes -

XX Claim 27; SEQ ID NO 12085; 327pp + sequence listing; English.  
 XX The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human breast and BT 474 cells. The method involves contacting  
 CC the probes with a collection of detectably labelled nucleic acids  
 CC derived from mRNA of human breast, and then measuring the label  
 CC bound to each probe of the microarray. The probes are useful for  
 CC verifying the expression of regions of genomic DNA predicted to  
 CC encode proteins. They are useful for gene discovery, and for  
 CC determining predisposition and/or prognosing breast disease. Gene  
 CC expression analysis is useful for assessing the toxicity of chemical  
 CC agents on cells. The microarray of this invention presents a far greater  
 CC diversity of probes for measuring gene expression, with far less bias  
 CC than expressed sequence tag microarrays. The method is suitable for  
 CC rapid production of functional information from genomic sequence. The  
 CC present sequence is a peptide encoded by a single exon nucleic acid  
 CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 48 AA;

Query Match 63.7%; Score 51; DB 22; Length 48;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 BEGGGGGGG 15  
 |||||  
 Db 6 BEGGGGGGG 14

RESULT 15

AAM55071  
 ID AAM55071 standard; Protein; 48 AA.

AC AAM55071;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe encoded protein SEQ ID NO: 27176.

XX Human; brain expressed exon; gene expression analysis; probe;  
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 KW epilepsy; cancer.

XX Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human  
 PT brains -  
 XX Example 4; SEQ ID NO: 27176; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is a protein encoded by one of  
 CC the probes of the invention.

XX SQ Sequence 48 AA;

Query Match 63.7%; Score 51; DB 22; Length 48;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 BEGGGGGGG 15  
 |||||  
 Db 6 BEGGGGGGG 14

Search completed: December 30, 2003, 12:01:34  
 Job time : 42 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2003, 12:00:43 ; Search time 32 Seconds  
(without alignments)  
93.289 Million cell updates/sec

Title: US-09-807-610-11

Perfect score: 80

Sequence: 1 ALADYEEGGGGGE 15

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 724715 seqs, 199017464 residues

Total number of hits satisfying chosen parameters: 724715

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80	100.0	180	9	US-09-730-617-62
2	80	100.0	180	9	US-09-730-617-79
3	80	100.0	180	15	US-10-167-127-8
4	51	63.7	29	12	US-10-029-386-32712
5	51	63.7	48	9	US-09-864-761-42759
6	51	63.7	77	9	US-09-864-761-35634
7	51	63.7	258	12	US-10-322-892-2
8	50	62.5	977	10	US-09-949-192-23
9	50	62.5	977	12	US-10-190-115-34
10	50	62.5	977	12	US-10-424-233-21
11	48	60.0	325	11	US-09-934-455-138
12	48	60.0	325	12	US-10-225-068-246
13	48	60.0	325	12	US-10-302-267-62
14	48	60.0	940	9	US-09-815-242-10447
15	48	60.0	941	9	US-09-815-242-13818

16	47	58.8	340	10	US-09-900-237-22	Sequence 22, Appl
17	47	58.8	670	11	US-09-893-519A-32	Sequence 32, Appl
18	47	58.8	1094	12	US-10-209-059-18	Sequence 18, Appl
19	47	58.8	1094	12	US-10-160-719-26	Sequence 26, Appl
20	47	58.8	1094	12	US-10-160-719-46	Sequence 46, Appl
21	47	58.8	1165	10	US-09-900-237-8	Sequence 8, Appl
22	46	57.5	72	12	US-10-029-386-31411	Sequence 31411, A
23	46	57.5	170	9	US-09-864-761-39624	Sequence 39624, A
24	46	57.5	311	11	US-09-934-455-172	Sequence 172, App
25	46	57.5	311	12	US-10-302-267-18	Sequence 18, Appl
26	46	57.5	311	15	US-10-278-173-12	Sequence 12, Appl
27	46	57.5	485	12	US-10-029-386-32442	Sequence 32442, A
28	46	57.5	958	9	US-09-801-260-2	Sequence 2, Appl
29	46	57.5	958	12	US-10-145-586-41	Sequence 41, Appl
30	45.5	56.9	431	8	US-08-822-186-2	Sequence 2, Appl
31	45.5	56.9	431	8	US-08-937-755-2	Sequence 2, Appl
32	45.5	56.9	431	8	US-08-957-425-2	Sequence 2, Appl
33	45.5	56.9	431	8	US-08-260-675-17	Sequence 17, Appl
34	45.5	56.9	431	9	US-09-045-331-2	Sequence 2, Appl
35	45.5	56.9	431	9	US-09-828-607-2	Sequence 2, Appl
36	45.5	56.9	431	10	US-09-982-543A-10	Sequence 10, Appl
37	45.5	56.9	431	11	US-09-039-107-2	Sequence 2, Appl
38	45.5	56.9	431	11	US-09-798-518-1	Sequence 1, Appl
39	45.5	56.9	431	11	US-09-540-466-2	Sequence 2, Appl
40	45.5	56.9	431	11	US-09-423-943-2	Sequence 2, Appl
41	45.5	56.9	431	12	US-10-286-152A-8	Sequence 8, Appl
42	45.5	56.9	431	12	US-10-301-822-10	Sequence 10, Appl
43	45.5	56.9	431	12	US-09-012-846-2	Sequence 2, Appl
44	45.5	56.9	431	12	US-09-960-789-1	Sequence 1, Appl
45	45.5	56.9	431	12	US-10-295-027-518	Sequence 518, App

## ALIGNMENTS

### RESULT 1

US-09-730-617-62  
; Sequence 62, Application US/09730617  
; Patent No. US20020068279A1  
; GENERAL INFORMATION:  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Prayaga, Sudhirdas K  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Mezes, Peter S  
; TITLE OF INVENTION: No. US20020068279A1e1 Proteins and Nucleic Acids Encoding the Sam  
; FILE REFERENCE: 15966-609  
; CURRENT APPLICATION NUMBER: US/09730, 617  
; CURRENT FILING DATE: 2000-12-05  
; PRIOR APPLICATION NUMBER: 60/169, 056  
; PRIOR FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: 60/169, 886  
; PRIOR FILING DATE: 1999-12-09  
; PRIOR APPLICATION NUMBER: 60/169, 866  
; PRIOR FILING DATE: 1999-12-09  
; PRIOR APPLICATION NUMBER: 60/170, 252  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: 60/175, 740  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 62  
; LENGTH: 180  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-730-617-62

Query Match 100.0%; Score 80; DB 9; Length 180;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 ALADYEEGGGGGE 15

Db 2 ALADLYEGGGGGG 16  
|||||

## RESULT 2

US-09-730-617-79

; Sequence 79, Application US/09730617  
; Patent No. US20020068279A1  
; GENERAL INFORMATION:  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Prayaga, Sudhirdas K  
; APPLICANT: Shinkete, Richard A  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Mezei, Peter S  
; TITLE OF INVENTION: No. US20020068279A1el Proteins and Nucleic Acids Encoding the Sam  
; FILE REFERENCE: 15966-609  
; CURRENT APPLICATION NUMBER: US/09/730,617  
; CURRENT FILING DATE: 2000-12-05  
; PRIOR APPLICATION NUMBER: 60/169,056  
; PRIOR FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: 60/169,886  
; PRIOR FILING DATE: 1999-12-09  
; PRIOR APPLICATION NUMBER: 60/169,866  
; PRIOR FILING DATE: 1999-12-09  
; PRIOR APPLICATION NUMBER: 60/170,252  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: 60/175,740  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 79  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-730-617-79

Query Match 100.0%; Score 80; DB 9; Length 180;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ALADLYEGGGGGG 15  
Db 2 ALADLYEGGGGGG 16  
|||||

## RESULT 3

US-10-167-127-8

; Sequence 8, Application US/10167127  
; Publication No. US20030100031A1  
; GENERAL INFORMATION:  
; APPLICANT: DOW, STEVEN  
; APPLICANT: DUFF, GORDON W  
; TITLE OF INVENTION: INTEGRATIVE ASSAYS FOR MONITORING MOLECULAR ASSEMBLY  
; FILE REFERENCE: MSA-026.01 (20974-2601)  
; CURRENT APPLICATION NUMBER: US/10/167,127  
; CURRENT FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: 60/297,305  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-167-127-8

Query Match 100.0%; Score 80; DB 15; Length 180;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ALADLYEGGGGGG 15

Db 2 ALADLYEGGGGGG 16  
|||||

## RESULT 4

US-10-029-386-32712

; Sequence 32712, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Hanzel, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
; FILE REFERENCE: A60MICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 32712  
; LENGTH: 29  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC007759.1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7  
US-10-029-386-32712

Query Match 63.7%; Score 51; DB 12; Length 29;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 BEGGGGGGG 15  
Db 20 BEGGGGGGG 28  
|||||

## RESULT 5

US-09-864-761-42759

; Sequence 42759, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: A60MICA-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42759
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009664.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
; NAME/KEY: unsure
; LOCATION: 45
; US-09-864-761-42759

```

```

Query Match 63.7%; Score 51; DB 9; Length 48;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 7 BEGGGGGG 15
Db 6 BEGGGGGG 14

```

```

RESULT 6
US-09-864-761-35634

```

```

; Sequence 35634, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn. Sharron G.
; APPLICANT: Hanzel, David K.
; APPLICANT: Hanzel, David K.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35634
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010165.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; US-09-864-761-35634

```

```

Query Match 63.7%; Score 51; DB 9; Length 77;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 7 BEGGGGGG 15
Db 22 BEGGGGGG 30

```

```

RESULT 7

```

```

US-10-322-892-3
; Sequence 3, Application US/10322892
; Publication No. US20030171257A1
; GENERAL INFORMATION:
; APPLICANT: STIRBEL, ROBERT C.
; APPLICANT: SNEAD, MALCOLM L.
; APPLICANT: XU, JIMMY
; APPLICANT: VITETTA, ELLEN S.
; APPLICANT: WILK, PETER J.
; TITLE OF INVENTION: METHOD AND RELATED COMPOSITION EMPLOYING NANOSTRUCTURES
; FILE REFERENCE: W07-505
; CURRENT APPLICATION NUMBER: US/10/322,892
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: 60/342,894
; PRIOR FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-322-892-3

```

```

Query Match 63.7%; Score 51; DB 12; Length 258;

```

Best Local Similarity 81.8%; Pred. No. 18;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 4 DLYEGGGGGG 14  
DB 122 DNEEGGGGGG 132

RESULT 8  
US-09-949-192-23  
; Sequence 23, Application US/09949192  
; Patent No. US20020142292A1  
; GENERAL INFORMATION:  
; APPLICANT: Parham, Christi L.  
; APPLICANT: Gorman, Daniel L.  
; APPLICANT: Kurata, Hirokazu  
; APPLICANT: Arai, Naoko  
; APPLICANT: Sana, Theodore R.  
; APPLICANT: Matteson, Jeanine D.  
; APPLICANT: Murphy, Erin E.  
; APPLICANT: Savkoor, Chetan  
; APPLICANT: Grein, Jeffery  
; APPLICANT: Smith, Kathleen M.  
; APPLICANT: McClanahan, Terrill K.  
; TITLE OF INVENTION: MAMMALIAN GENES; RELATED REAGENTS AND METHODS  
; CURRENT APPLICATION NUMBER: US/09/949,192  
; CURRENT FILING DATE: 2001-09-07  
; PRIOR APPLICATION NUMBER: 60/231,267  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 23  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-949-192-23

Query Match 62.5%; Score 50; DB 10; Length 977;  
Best Local Similarity 80.0%; Pred. No. 87;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 5 LYEGGGGGG 14  
DB 708 LFDGGGGGGG 717

RESULT 9  
US-10-190-115-34  
; Sequence 34, Application US/10190115  
; Publication No. US20030207394A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsbrook, John P. II  
; APPLICANT: Boldog, Ferenc L.  
; APPLICANT: Burgess, Catherine E.  
; APPLICANT: Casman, Stacie J.  
; APPLICANT: Grose, William M.  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Ji, Weizhen  
; APPLICANT: Lepley, Denise M.  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Mezick, Amanda J.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Shen, Lei  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Spaderna, Steven K.  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Szekeres, Edward S. Jr.  
; APPLICANT: Taupier, Raymond J. Jr.  
; APPLICANT: Tchernev, Velizar T.

; APPLICANT: Zethusen, Bryan D.  
; APPLICANT: Voss, Edward Z.  
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
; FILE REFERENCE: 21402-050 CIP  
; CURRENT APPLICATION NUMBER: US/10/190,115  
; CURRENT FILING DATE: 2003-02-10  
; PRIOR APPLICATION NUMBER: 60/303,168  
; PRIOR FILING DATE: 2001-07-05  
; PRIOR APPLICATION NUMBER: 60/368,996  
; PRIOR FILING DATE: 2002-04-01  
; PRIOR APPLICATION NUMBER: 60/386,816  
; PRIOR FILING DATE: 2002-06-07  
; PRIOR APPLICATION NUMBER: 60/215,854  
; PRIOR FILING DATE: 2000-07-03  
; PRIOR APPLICATION NUMBER: 60/215,856  
; PRIOR FILING DATE: 2000-07-03  
; PRIOR APPLICATION NUMBER: 60/215,902  
; PRIOR FILING DATE: 2000-07-03  
; PRIOR APPLICATION NUMBER: 60/216,585,  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/216,586  
; PRIOR FILING DATE: 2001-07-07  
; PRIOR APPLICATION NUMBER: 60/216,722  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/219,622  
; PRIOR FILING DATE: 2000-07-17  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 136  
; SOFTWARE: Curaseqlist version 0.1  
; SEQ ID NO 34  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-190-115-34

Query Match 62.5%; Score 50; DB 12; Length 977;  
Best Local Similarity 80.0%; Pred. No. 87;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 5 LYEGGGGGG 14  
DB 708 LFDGGGGGGG 717

RESULT 10  
US-10-424-233-21  
; Sequence 21, Application US/10424233  
; Publication No. US2003020263A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: NOVEL HUMAN LEUCINE-RICH REPEAT-CONTAINING PROTEINS SPECIFICALLY  
; FILE REFERENCE: D0233 NP  
; CURRENT APPLICATION NUMBER: US/10/424,233  
; CURRENT FILING DATE: 2003-04-25  
; PRIOR APPLICATION NUMBER: U.S. 60/375,335  
; PRIOR FILING DATE: 2002-04-25  
; NUMBER OF SEQ ID NOS: 75  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 21  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-424-233-21

Query Match 62.5%; Score 50; DB 12; Length 977;  
Best Local Similarity 80.0%; Pred. No. 87;  
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 5 LYEGGGGGG 14  
DB 708 LFDGGGGGGG 717

RESULT 11  
US-09-934-455-138  
; Sequence 138, Application US/09934455  
; Publication No. US20030121070A1  
; GENERAL INFORMATION:  
; APPLICANT: Adam, Luc  
; APPLICANT: Creelman, Robert  
; APPLICANT: Dubell, Arnold  
; APPLICANT: Heard, Jacqueline  
; APPLICANT: Jiang, Cai-Zhong  
; APPLICANT: Keddie, James  
; APPLICANT: Pilgrim, Marsha  
; APPLICANT: Ratcliffe, Oliver  
; APPLICANT: Reuber, Lynne  
; APPLICANT: Riechmann, Jose Luis  
; APPLICANT: Yu, Guo-Liang  
; APPLICANT: Pineda, Omaira  
; FILE OF INVENTION: Genes for Modifying Plant Traits IV  
; FILE REFERENCE: MBI-0025  
; CURRENT APPLICATION NUMBER: US/09/934,455  
; CURRENT FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: 60/227439  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: MBI-0022  
; PRIOR FILING DATE: 2001-11-16  
; PRIOR APPLICATION NUMBER: MBI-0023  
; PRIOR FILING DATE: 2001-04-17  
; NUMBER OF SEQ ID NOS: 516  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 138  
; LENGTH: 325  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-934-455-138

Query Match 60.0%; Score 48; DB 11; Length 325;  
Best Local Similarity 71.4%; Pred. No. 56;  
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 LADLYEEGGGGGGE 15  
||| |||||  
DB 264 LASGYGGGGGGGGE 277

RESULT 12  
US-10-225-068-246  
; Sequence 246, Application US/10225068  
; Publication No. US20030217383A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendel Biotechnology, Inc.  
; APPLICANT: Reuber, T. Lynne  
; APPLICANT: Riechmann, Jose Luis  
; APPLICANT: Heard, Jacqueline E.  
; APPLICANT: Jiang, Cai-Zhong  
; APPLICANT: Adam, Luc J.  
; APPLICANT: Dubell, Arnold T.  
; APPLICANT: Ratcliffe, Oliver  
; APPLICANT: Pineda, Omaira  
; APPLICANT: Yu, Guo-Liang  
; APPLICANT: Broun, Pierre E.  
; FILE OF INVENTION: STRESS-RELATED POLYNUCLEOTIDES AND  
; FILE REFERENCE: 51442002040  
; CURRENT APPLICATION NUMBER: US/10/225,068  
; CURRENT FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: 60/310,847  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: 60/336,049  
; PRIOR FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: 60/338,692  
; PRIOR FILING DATE: 2001-12-11  
; PRIOR APPLICATION NUMBER: 10/171,468

US-09-807-610-11.rapb

PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 246  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 246  
; LENGTH: 325  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (51)...(114)  
; OTHER INFORMATION: Conserved domain  
US-10-225-068-246

Query Match 60.0%; Score 48; DB 12; Length 325;  
Best Local Similarity 71.4%; Pred. No. 56;  
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 LADLYEEGGGGGGE 15  
||| |||||  
DB 264 LASGYGGGGGGGGE 277

RESULT 13  
US-10-302-267-62  
; Sequence 62, Application US/10302267  
; Publication No. US20030229915A1  
; GENERAL INFORMATION:  
; APPLICANT: Keddie, James  
; APPLICANT: Fromm, Michael  
; APPLICANT: Heard, Jacqueline  
; APPLICANT: Riechmann, Jose Luis  
; APPLICANT: Adam, Luc  
; APPLICANT: Broun, Pierre  
; APPLICANT: Pineda, Omaira  
; APPLICANT: Reuber, Lynne  
; APPLICANT: Zhang, James  
; APPLICANT: Yu, Guo-Liang  
; APPLICANT: Jiang, Cai-Zhong  
; APPLICANT: Samaha, Raymond  
; APPLICANT: Pilgrim, Marsha  
; APPLICANT: Creelman, Robert  
; TITLE OF INVENTION: PLANT GENE SEQUENCES II  
; FILE REFERENCE: MBI-0007  
; CURRENT APPLICATION NUMBER: US/10/302,267  
; CURRENT FILING DATE: 2002-11-22  
; PRIOR APPLICATION NUMBER: US/09/506,720  
; PRIOR FILING DATE: 2000-02-17  
; PRIOR APPLICATION NUMBER: 60/120,880  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: 60/121,037  
; PRIOR FILING DATE: 1999-02-22  
; PRIOR APPLICATION NUMBER: 60/124,278  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: 60/129,450  
; PRIOR FILING DATE: 1999-04-15  
; PRIOR APPLICATION NUMBER: 60/135,134  
; PRIOR FILING DATE: 1999-05-20  
; PRIOR APPLICATION NUMBER: 60/144,153  
; PRIOR FILING DATE: 1999-07-15  
; PRIOR APPLICATION NUMBER: 60/161,143  
; PRIOR FILING DATE: 1999-10-22  
; PRIOR APPLICATION NUMBER: 60/162,656  
; NUMBER OF SEQ ID NOS: 218  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 62  
; LENGTH: 325  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: G805  
US-10-302-267-62

Query Match 60.0%; Score 48; DB 12; Length 325;  
Best Local Similarity 71.4%; Pred. No. 56;  
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 LADLYEFGGGGG 15  
Db 264 LASGYGGGGGG 277

## RESULT 14

US-09-815-242-10447  
; Sequence 10447, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10447  
; LENGTH: 940  
; TYPE: PRT

; ORGANISM: Escherichia coli  
US-09-815-242-10447

Query Match 60.0%; Score 48; DB 9; Length 940;  
Best Local Similarity 64.3%; Pred. No. 1.5e+02;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 LADLYEFGGGGG 15  
Db 901 IVDLFGGGGGGG 914

## RESULT 15

US-09-815-242-13818  
; Sequence 13818, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13818  
; LENGTH: 941  
; TYPE: PRT  
; ORGANISM: Salmonella typhi  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(941)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-815-242-13818

Query Match 60.0%; Score 48; DB 9; Length 941;  
Best Local Similarity 64.3%; Pred. No. 1.5e+02;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 LADLYEFGGGGG 15  
Db 901 IVDLFGGGGGGG 914

Search completed: December 30, 2003, 12:04:04  
Job time : 32 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2003, 12:00:03 ; Search time 21 Seconds  
(without alignments)  
68.692 Million cell updates/sec

Title: US-09-807-610-11

Perfect score: 80  
Sequence: 1 ALADLYEEGGGGGE 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first: 45 summaries

Database: PIR 76:  
1: Pirl:  
2: Pirl2:  
3: Pirl3:  
4: Pirl4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	100.0	180	2 A39386	interleukin-1 rece
2	49	61.3	206	2 T32854	hypothetical prote
3	48	60.0	325	2 G96718	unknown protein, 5
4	48	60.0	940	1 BVECUA	excinnuclease ABC c
5	48	60.0	940	2 H91258	excision nuclease
6	48	60.0	940	2 D86099	excision nuclease
7	48	60.0	941	2 AB1017	excision nuclease
8	48	60.0	947	2 AF0040	excinnuclease ABC c
9	47	58.8	201	2 J01094	hypothetical 20.2K
10	47	58.8	726	2 G83310	conserved hypotet
11	46.5	58.1	158	2 F96632	hypothetical prote
12	46.5	58.1	455	2 B86427	hypothetical prote
13	46	57.5	269	2 A70719	probable enoyl-coA
14	46	57.5	311	2 F86341	hypothetical prote
15	46	57.5	385	2 T20410	hypothetical prote
16	46	57.5	578	2 T48795	hypothetical prote
17	46	57.5	581	2 T22341	origin recognition
18	46	57.5	750	2 T10864	transcription acti
19	45.5	56.9	431	1 BMH07	bone morphogenetic
20	45.5	56.9	1171	2 F83110	exodeoxyribonuclea
21	45	56.2	386	1 S22315	snRNP-associated p
22	45	56.2	396	2 S56820	hypothetical prote
23	45	56.2	404	2 S13648	SCJ1 protein - yea
24	45	56.2	404	2 S54729	RNA-binding protei
25	45	56.2	629	2 F84428	probable myosin he
26	45	56.2	653	2 C84648	hypothetical prote
27	45	56.2	856	2 G96814	hypothetical prote
28	45	56.2	943	2 D64057	excinnuclease ABC c
29	45	56.2	956	2 AB1761	excinnuclease ABC (

## ALIGNMENTS

### RESULT 1

A39386  
interleukin-1 receptor antagonist, long intracellular splice form - human  
N:Contains: interleukin-1 receptor antagonist, short intracellular splice form  
C:Species: Homo sapiens (man)  
C>Date: 28-Feb-1992 #sequence\_revision 11-Apr-1997 #text\_change 26-May-2000  
C:Accession: I37893; A39386  
R:Muzio, M.; Folentaru, N.; Sironi, M.; Poli, G.; De Gioia, L.; Introna, M.; Mantovan, J. Exp. Med. 182, 623-628, 1995  
A:Title: Cloning and characterization of a new isoform of the interleukin 1 receptor ant:  
A:Reference number: I37893; MUID:95355865; PMID:7629520  
A:Accession: I37893  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-180 <RES>  
A:Cross-references: EMBL:X84348; NID:g1008970; PIDN:CAA59087.1; PID:g1008971  
R:Hasikill, S.; Martin, G.; Van Le, L.; Morris, J.; Peace, A.; Bigler, C.F.; Jaffe, G.J.; Proc. Natl. Acad. Sci. U.S.A. 88, 3681-3685, 1991  
A:Title: cDNA cloning of an intracellular form of the human interleukin 1 receptor antag:  
A:Reference number: A39386; MUID:91219436; PMID:1827201  
A:Accession: A39386  
A:Molecule type: mRNA  
A:Residues: 1-3,25-180 <HAS>  
A:Cross-references: GB:M55646; NID:g186291; PIDN:AAA59138.1; PID:g186292  
C:Comment: For an alternative splice form, see PIR:A30368  
C:Genetics:  
A:Gene: GDB:IL1RN  
A:Cross-references: GDB:125897; OMIM:147679  
A:Map position: 2q14.2-2q14.2  
C:Superfamily: interleukin-1  
C:Keywords: alternative splicing; cytokine receptor  
F:1-180/Product: interleukin-1 receptor antagonist, long intracellular splice form #statu  
F:1-3,25-180/Product: interleukin-1 receptor antagonist, short intracellular splice form

Query Match 100.0%; Score 80; DB 2; Length 180;

Best Local Similarity 100.0%; Pred. No. 0.00015;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ALADLYEEGGGGGE 15

Db 2 ALADLYEEGGGGGE 16

### RESULT 2

T32854

hypothetical protein Y8G1A.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T32854

R:Cordes, M.; Le, T.T.

submitted to the EMBL Data Library, December 1997

A:Description: The sequence of C. elegans cosmid Y8G1A.

A:Reference number: Z21234  
 A:Accession: T32854  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-206 <COR>  
 A:Cross-references: EMBL:AF040656; PIDN:AB95048.1; GSPDB:GN00019; CESP:Y8G1A.1  
 A:Experimental source: strain Bristol N2; clone Y8G1A  
 C:Genetics:  
 A:Gene: CESP:Y8G1A.1  
 A:Map position: 1  
 A:Introns: 26/1, 47/1, 107/3

Query Match 61.3%; Score 49; DB 2; Length 206;  
 Best Local Similarity 63.6%; Pred. No. 5.3;  
 Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 4 DLYREGGGGGG 14  
 |:::|||||  
 Db 156 DIFKGGGGGG 166

RESULT 3  
 G96718  
 unknown protein, 54453-53476 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: G96718  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;  
 ansen, N.P.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000.  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: G96718  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-325 <STO>  
 A:Cross-references: GB:AB005173; NID:g6665547; PIDN:AAF22916.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: T6C23.11  
 A:Map position: 1

Query Match 60.0%; Score 48; DB 2; Length 325;  
 Best Local Similarity 71.4%; Pred. No. 12;  
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 LADLYREGGGGGG 15  
 |||::|  
 Db 264 LASYGGGGGGGG 277

RESULT 4  
 BVBCUA  
 excinuclease ABC chain A - Escherichia coli (strain K-12)  
 N:Alternate names: uvrA protein  
 N:Contains: excision endonuclease ABC (EC 3.1.-.-) chain A  
 C:Species: Escherichia coli  
 C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 01-Mar-2002  
 C:Accession: A23869; I78011; A65214  
 R:Husain, I.; Van Houten, B.; Thomas, D.C.; Sancar, A.  
 J. Biol. Chem. 261, 4895-4901, 1986  
 A:Title: Sequences of Escherichia coli uvra gene and protein reveal two potential ATP b  
 A:Reference number: A23869; MUID:86168204; PMID:3007478  
 A:Accession: A23869  
 A:Molecule type: DNA  
 A:Residues: 1-940 <HUS>  
 A:Cross-references: GB:M13495; NID:g148164; PIDN:AAA24754.1; PID:g148165  
 R:Sancar, A.; Sancar, G.B.; Rupp, W.D.; Little, J.W.; Mount, D.W.

Nature 298, 96-98, 1982  
 A:Title: LexA protein inhibits transcription of the E. coli uvra gene in vitro.  
 A:Reference number: I58044; MUID:82220077; PMID:6283374  
 A:Accession: I78011  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-14 <RES>  
 A:Cross-references: GB:J01721; NID:g148161; PIDN:AAA24753.1; PID:g551847  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617; PMID:9278503  
 A:Accession: A65214  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-940 <BLAT>  
 A:Cross-references: GB:AB000479; GB:U00096; NID:g2367340; PIDN:AACT7028.1; PID:g2367343;  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Comment: This protein is an ATPase and a DNA-binding protein that preferentially binds  
 e, an ATP-dependent DNA repair enzyme that catalyzes the excision reaction of UV-damaged  
 C:Genetics:  
 A:Gene: uvra  
 A:Map position: 92 min  
 C:Superfamily: excinuclease ABC chain A; ATP-binding cassette homology  
 C:Keywords: ATP; DNA binding; DNA repair; duplication; hydrolase; nucleotide binding; P-1  
 F:31-38/Region: nucleotide-binding motif A (P-loop)  
 F:623-907/Domain: ATP-binding cassette homology <ABCE>  
 F:640-647/Region: nucleotide-binding motif A (P-loop)

Query Match 60.0%; Score 48; DB 1; Length 940;  
 Best Local Similarity 64.3%; Pred. No. 34;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 LADLYREGGGGGG 15  
 :|||::|  
 Db 901 IVDLPGGGGGGG 914

RESULT 5  
 H91258  
 excinuclease subunit A [imported] - Escherichia coli (strain O157:H7, substrain RIM  
 C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 C:Accession: H91258  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genou  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: H91258  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-940 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA838463.1; PID:g13364517; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RIMD 0509952  
 C:Genetics:  
 A:Gene: EC5040  
 C:Superfamily: excinuclease ABC chain A; ATP-binding cassette homology

Query Match 60.0%; Score 48; DB 2; Length 940;  
 Best Local Similarity 64.3%; Pred. No. 34;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 LADLYREGGGGGG 15  
 :|||::|  
 Db 901 IVDLPGGGGGGG 914

RESULT 6  
 D86099  
 excinuclease subunit A [imported] - Escherichia coli (strain O157:H7, substrain EDL5  
 C:Species: Escherichia coli



C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: D86099  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grotbeck, E.J.; Davis, A.; Lim, A.; Dimalanta, E.; Potamousta, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A/Reference number: A85480; MUID:21074935; PMID:11206551  
 A/Accession: D86099  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-940 <STO>  
 A/Cross-references: GB:AE005174; NID:g12519009; PIDN:AAG59256.1; GSPDB:GN00145; UWGP:256  
 A/Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: *uvrA*  
 C:Superfamily: excinuclease ABC chain A; ATP-binding cassette homology

Query Match 60.0%; Score 48; DB 2; Length 940;  
 Best Local Similarity 64.3%; Pred. No. 34;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 LADLYEKGCGGCGE 15  
 : || ||| ||||  
 Db 901 IVDLGPFGGCGGGE 914

RESULT 7  
 AE1017  
 excinuclease chain A [imported] - Salmonella enterica subsp. enterica serovar Typhi  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A/Note: this species has also been called Salmonella typhi  
 C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C:Accession: AE1017  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Connor, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
 A/Reference number: AB0502; MUID:21534947; PMID:11677608  
 A/Accession: AE1017  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-941 <PAR>  
 A/Cross-references: GB:AL513382; PIDN:CAD09238.1; PID:g16505242; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: *uvrA*  
 C:Superfamily: excinuclease ABC chain A; ATP-binding cassette homology

Query Match 60.0%; Score 48; DB 2; Length 941;  
 Best Local Similarity 64.3%; Pred. No. 34;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 LADLYEKGCGGCGE 15  
 : || ||| ||||  
 Db 901 IVDLGPFGGCGGGE 914

RESULT 8  
 AF0040  
 excinuclease ABC chain A *uvrA* [imported] - Yersinia pestis (strain C092)  
 C:Species: Yersinia pestis  
 C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
 C:Accession: AF0040  
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.  
 deno-Tarraga, A.; Wren, B.W.; Chillingworth, T.; Cronin, A.; Davies, P.; Dougan, G.;  
 ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,  
 Nature 413, 523-527, 2001  
 A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A/Reference number: AB0001; MUID:21470413; PMID:11586360  
 A/Accession: AF0040  
 A/Status: preliminary  
 A/Molecule type: DNA

A/Residues: 1-947 <KUR>  
 A/Cross-references: GB:AL590842; PIDN:CAC89185.1; PID:g15978424; GSPDB:GN00175  
 C:Genetics:  
 A:Gene: *uvrA*  
 C:Superfamily: excinuclease ABC chain A; ATP-binding cassette homology

Query Match 60.0%; Score 48; DB 2; Length 947;  
 Best Local Similarity 64.3%; Pred. No. 34;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 LADLYEKGCGGCGE 15  
 : || ||| ||||  
 Db 901 IVDLGPFGGCGGGE 914

RESULT 9  
 JQ1094  
 hypothetical 20.2K protein - tomato ringspot virus  
 C:Species: tomato ringspot virus  
 C>Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 08-Oct-1999  
 C:Accession: JQ1094  
 R:Rott, M.E.; Tremaine, J.H.; Rochon, D.M.  
 J. Gen. Virol. 72, 1505-1514, 1991  
 A/Title: Nucleotide sequence of tomato ringspot virus RNA-2.  
 A/Reference number: JQ1093; MUID:91311402; PMID:1856689  
 A/Accession: JQ1094  
 A/Status: translation not shown  
 A/Molecule type: genomic RNA  
 A/Residues: 1-201 <ROT>  
 A/Cross-references: GB:D12477; GB:D01129; NID:g2222674; PIDN:BAA02044.1; PID:d1002536; PI  
 A/Experimental source: strain raspberry

Query Match 58.8%; Score 47; DB 2; Length 201;  
 Best Local Similarity 75.0%; Pred. No. 10;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 ADLYEKGCGGCGG 14  
 : || ||| ||||  
 Db 55 ANLAAGGCGGCGG 66

RESULT 10  
 G83310  
 conserved hypothetical protein PA2685 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: G83310  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam,  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A/Reference number: AB2950; MUID:20437337; PMID:10984043  
 A/Accession: G83310  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-726 <STO>  
 A/Cross-references: GB:AE004696; GB:AE004091; NID:g9948750; PIDN:AAG06073.1; GSPDB:GN001  
 A/Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA2685

Query Match 58.8%; Score 47; DB 2; Length 726;  
 Best Local Similarity 72.7%; Pred. No. 36;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 DLYEKGCGGCGG 14  
 : ||| ||| ||||  
 Db 367 ELYETGCGGCGG 377

RESULT 11  
 F96692

hypothetical protein T4024.4 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: F96692  
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzborg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: F96692  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-158 <STO>  
 A:Cross-references: GB:AE005173; NID:g11128395; PIDN:AG31200.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: T4024.4  
 A:Map position: 1

Query Match 58.1%; Score 46.5; DB 2; Length 158;  
 Best Local Similarity 47.6%; Pred. No. 9.3;  
 Matches 10; Conservative 2; Mismatches 2; Indels 7; Gaps 1;

QY 2 LADLY-----EEGGGGGG 15

DB 72 LGDLFVSPFPFEEGGGGGG 92

RESULT 12

B86427  
 hypothetical protein F12P21.7 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 27-Nov-2001  
 C:Accession: B86427  
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzborg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: B86427  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-455 <STO>  
 A:Cross-references: GB:AE005172; NID:g11120781; PIDN:AG30962.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1

Query Match 58.1%; Score 46.5; DB 2; Length 455;  
 Best Local Similarity 83.3%; Pred. No. 27;  
 Matches 10; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 4 DLY-BEGGGGGG 14

DB 295 DRYEBEGGGGGG 306

RESULT 13

A70719  
 probable enoyl-CoA hydratase - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C:Accession: A70719

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
 A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987; PMID:9634230  
 A:Accession: A70719  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-269 <COL>  
 A:Cross-references: GB:Z79700; GB:AL123456; NID:g3261628; PIDN:CAB02007.1; PID:g1524191  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: echA7  
 C:Superfamily: naphthoate synthase; enoyl-CoA hydratase homology  
 F:29-183/Domain: enoyl-CoA hydratase homology <ECH>

Query Match 57.5%; Score 46; DB 2; Length 269;  
 Best Local Similarity 81.8%; Pred. No. 19;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ADLYBEGGGGG 13

DB 68 ADLSEAGGGGG 78

RESULT 14

F86341  
 hypothetical protein F9H16.12 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
 C:Accession: F86341  
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzborg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: F86341  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-311 <STO>  
 A:Cross-references: GB:AE005172; NID:g4836899; PIDN:AAD30602.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 57.5%; Score 46; DB 2; Length 311;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 BEGGGGGG 14

DB 249 BEGGGGGG 256

RESULT 15

T20410  
 hypothetical protein E02A10.2 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T20410  
 R:Thomas K.  
 submitted to the EMBL Data Library, October 1996  
 A:Reference number: Z19271  
 A:Accession: T20410  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA  
A;Residues: 1-385 <WIL>  
A;Cross-references: EMBL:Z01053; PIDN:CAB02877.1; GSPDB:GN00023; CESP:E02A10.2  
A;Experimental source: clone E02A10  
C;Genetics:  
A;Gene: CESP:E02A10.2  
A;Map position: 5  
A;Introns: 32/1; 72/1; 85/1; 122/1; 133/1; 220/3  
Query Match 57.5%; Score 46; DB 2; Length 385;  
Best Local Similarity 64.3%; Pred. No. 27;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 1 ALADLYEEGGGGG 14  
Db 15 ATAFLPSSGGGGG 28

Search completed: December 30, 2003, 12:03:20  
Job time : 22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2003, 11:56:23 ; Search time 11 Seconds  
(without alignments)  
64.127 Million cell updates/sec

Title: US-09-807-610-11

Perfect score: 80

Sequence: 1. ALADLYEKGGGGG 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	62.5	977	1 Y848 HUMAN	O94933 homo sapien
2	48	60.0	576	1 Z384 HUMAN	O8tf68 homo sapien
3	48	60.0	579	1 Z384 RAT	O9eqj4 rattus norv
4	48	60.0	940	1 UVRA_ECO57	O8X5u9 escherichia
5	48	60.0	940	1 UVRA_ECO57	O8f602 escherichia
6	48	60.0	940	1 UVRA_ECOLI	P07671 escherichia
7	48	60.0	941	1 UVRA_SALTY	P37434 salmonella
8	48	60.0	941	1 UVRA_SALTY	O8z107 versinia pe
9	47	58.8	201	1 YR21_TESVR	P25245 tomato ring
10	46	57.5	311	1 ESCA_ARATH	O987c9 arabidopsis
11	46	57.5	1175	1 HCN4_RABIT	O9tv66 oryctolagus
12	45.5	56.9	431	1 BME7_HUMAN	P18075 homo sapien
13	45	56.2	386	1 RB87_DROME	P48810 drosophila
14	45	56.2	396	1 YJ88_YEAST	P47049 saccharomyc
15	45	56.2	404	1 CAZ_DROME	O27294 drosophila
16	45	56.2	404	1 SCUI_YEAST	P25303 saccharomyc
17	45	56.2	490	1 JIPI_DROME	P9w0k0 drosophila
18	45	56.2	943	1 UVRA_HAETN	P44410 haemophilus
19	45	56.2	973	1 UVRA_RHILO	O98m36 rhizobium 1
20	44	55.0	199	1 AC22_TENNO	P28968 tenebrio mo
21	44	55.0	342	1 ROAI_SCHAM	P21522 schistocerc
22	44	55.0	358	1 CBEA_HUMAN	P49715 homo sapien
23	44	55.0	377	1 GP27 RAT	O91jh3 rattus norv
24	44	55.0	601	1 HR78_DROME	O24142 drosophila
25	44	55.0	943	1 UVRA_PASMU	P57979 pasteurella
26	43	53.8	305	1 ROAO_HUMAN	O13151 homo sapien
27	43	53.8	375	1 GP27_HUMAN	O9ns67 homo sapien
28	43	53.8	460	1 SR54_HALNI	O9hnm5 halobacteri
29	43	53.8	502	1 XYLEB_LACBR	P35850 lactobacilli
30	43	53.8	515	1 PVRI_MOUSE	O91kff6 mus musculu
31	43	53.8	517	1 PVRI_HUMAN	O15223 homo sapien
32	43	53.8	642	1 HS72_PICAN	P33623 pichia angu
33	43	53.8	687	1 WRK2_ARATH	O9fg77 arabidopsis

34	43	53.8	862	1 CNRC_CHICK	P52731 gallus gall
35	43	53.8	942	1 UVRA_STRPY	O99y84 streptococc
36	43	53.8	950	1 UVRA_NEIGO	O50968 neisseria g
37	43	53.8	957	1 UVRA_BACSU	O34863 bacillus eu
38	43	53.8	1136	1 PHY_FICAB	O40762 picea abies
39	43	53.8	1191	1 S122_SQUAC	P55013 squaleus aca
40	43	53.8	1407	1 CYAA_DICDI	O03100 dictyostell
41	42.5	53.1	471	1 PPAP_CABEL	O09448 caenorhabdi
42	42	52.5	118	1 CRYP_CRYPA	P52753 cryphonectr
43	42	52.5	165	1 GRP1_ORYZA	P25074 oryza sativ
44	42	52.5	256	1 TRN9_MOUSE	P20334 mus musculu
45	42	52.5	280	1 LEX1_MOUSE	P52355 mus musculu

## ALIGNMENTS

RESULT 1  
Y848 HUMAN  
ID Y848 HUMAN STANDARD; PRT; 977 AA.  
AC O94933;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein KIAA0848.  
GN KIAA0848.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=99156230; PubMed=10048485;  
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro."  
RL DNA Res. 5:355-364(1998).  
CC - SIMILARITY: Contains 12 leucine-rich (LRR) repeats.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AB020655; BAA74871.1; --  
CC InterPro; IPR001611; LRR.  
CC InterPro; IPR000483; LRR\_Cterm.  
CC InterPro; IPR000372; LRR\_Nterm.  
CC InterPro; IPR003591; LRR\_Typ.  
CC Pfam; PF00560; LRR; 7.  
CC SMART; SM00082; LRRCT; 2.  
CC SMART; SM00013; LRRNT; 1.  
CC Hypothetical protein; Repeat; Leucine-rich repeat.  
CC REPEAT 76 99  
CC REPEAT 100 123  
CC REPEAT 125 147  
CC REPEAT 148 171  
CC REPEAT 172 195  
CC REPEAT 197 219  
CC REPEAT 407 430  
CC REPEAT 431 454  
CC REPEAT 455 478  
CC REPEAT 480 502  
CC REPEAT 503 526  
CC REPEAT 520 550

```

FT DOMAIN 712 723 POLY-GLY.
SQ SEQUENCE 977 AA; 109005 MW; 3C936B7E0003DF54 CRC64;

Query Match 62.5%; Score 50; DB 1; Length 977;
Best Local Similarity 80.0%; Pred. No. 12;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 5 LYREGGGGGG 14
DB 708 LFEDGGGGGG 717

RESULT 2
ID Z384 HUMAN STANDARD; PRT; 576 AA.
AC Q8TF68; O15407; Q8N938;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE zinc finger protein 384 (Nuclear matrix transcription factor 4)
DE (CAG repeat protein 1).
GN ZNF384 OR NMP4 OR CAGH1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Matsuo M.Y.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RA TISSUE=Brain;
RA Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M., Shimizu F.,
RA Wakebe H., Ono T., Hishigaki H., Watanabe T., Ozaki K., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Murakawa K., Kaneshori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K.,
RA Masuno Y., Nagai K., Isogai T.;
RA "NEBO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 395-576 FROM N.A.
RC TISSUE=Brain cortex;
RX MEDLINE=97369492; PubMed=9225980;
RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.-H., Kidwai A.S.,
RA Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
RT "CDNAs with long CAG trinucleotide repeats from human brain.";
RL Hum. Genet. 100:114-122(1997).
CC -1- FUNCTION: Transcription factor that binds the consensus DNA
CC sequence [GC]AATAA. Seems to bind and regulate the promoters of
CC MZF1, MZF3, MZF7 and COL1A1 (By similarity).
CC -1- SUBUNIT: Interacts with Cas (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q8TF68-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8TF68-2; Sequence=VSP_006920;
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: Contains 8 C2H2-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

or send an email to license@isb-sib.ch).
CC EMBL; AB070238; BAB85125.1;
CC EMBL; AK095734; BAC04618.1;
CC EMBL; U00738; AAB91437.1;
CC Genew; HGNC:11955; ZNF384.
CC InterPro; IPR007087; Znf_C2H2.
CC Pfam; PF00096; zf-C2H2; 8.
CC ProDom; PD000003; Znf_C2H2; 4.
CC SMART; SM00355; Znf_C2H2; 8.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 8.
CC Transcription regulation; zinc-finger; Metal-binding; Nuclear protein;
CC DNA-binding; Repeat; Alternative splicing.
CC ZN_FING 228 250
CC ZN_FING 256 278
CC ZN_FING 284 306
CC ZN_FING 317 339
CC ZN_FING 345 367
CC ZN_FING 373 397
CC ZN_FING 403 425
CC ZN_FING 433 455
CC ZN_FING 461 521
CC DOMAIN 466 499
CC DOMAIN 300 360
CC VARSPLIC
CC SEQUENCE 576 AA; 63091 MW; 2A152786C3C46D90 CRC64;
/FTId=VSP_006920
Query Match 60.0%; Score 48; DB 1; Length 576;
Best Local Similarity 90.0%; Pred. No. 13;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 LYREGGGGGG 14
DB 172 LYREGGGGGG 181

RESULT 3
ID Z384 RAT STANDARD; PRT; 579 AA.
AC Q9EQJ4; Q9EQJ2; Q9EQJ3; Q9JMJ5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 384 (Nuclear matrix transcription factor 4)
DE (Cas-associated zinc finger protein).
GN ZNF384 OR NMP4 OR CIZ.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), SUBCELLULAR LOCATION, DNA-BINDING, AND
RP INTERACTION WITH CAS.
RX MEDLINE=20136045; PubMed=10669742;
RA Nakamoto T., Yamagata T., Sakai R., Ogawa S., Honda H., Ueno H.,
RA Hirano N., Yazaki Y., Hirai H.;
RT "Ciz, a zinc finger protein that interacts with p130cas and activates
RT the expression of matrix metalloproteinases.";
RL Mol. Cell. Biol. 20:1649-1658(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC STEIN-Sprague-Dawley;
RX MEDLINE=21024193; PubMed=11149472;
RA Thunyakitpisal P., Alvarez M., Tokunaga K., Onyia J.B., Hock J.,
RA Ohashi N., Feister H., Rhodes S.J., Bidwell J.P.;
RT "Cloning and functional analysis of a family of nuclear matrix
RT transcription factors (NP/NMP4) that regulate type I collagen
RT expression in osteoblasts.";
RL J. Bone Miner. Res. 16:10-23(2001).
CC -1- FUNCTION: Transcription factor that binds the consensus DNA
CC sequence [GC]AATAA. Seems to bind and regulate the promoters of
CC MMP1, MMP3, MMP7 and COL1A1.

```

CC -1- SUBUNIT: Interacts with Cas.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=3;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=1;  
 CC IsoId=Q9EQJ4-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q9EQJ4-2; Sequence=VSP\_006921;  
 CC Name=3;  
 CC IsoId=Q9EQJ4-3; Sequence=VSP\_006922;  
 CC -1- TISSUE SPECIFICITY: Expressed in osteocytes, osteoblasts, and  
 CC chondrocytes in bone.  
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.  
 CC -1- SIMILARITY: Contains 8 C2H2-type zinc fingers.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AB019281; BAAG9664.1; -  
 DR EMBL; AF216804; AAG40582.1; -  
 DR EMBL; AF216805; AAG40583.1; -  
 DR EMBL; AF216806; AAG40584.1; -  
 DR HSSP; P08153; 1ZFD.  
 DR TRANSFAC; T05136; -  
 DR TRANSFAC; T05137; -  
 DR TRANSFAC; T05138; -  
 DR TRANSFAC; T05141; -  
 DR TRANSFAC; T05142; -  
 DR InterPro; IPR007087; Znf C2H2.  
 DR Pfam; PF00096; Znf C2H2; 8.  
 DR ProDom; PD000003; Znf C2H2; 4.  
 DR SMART; SM00355; Znf C2H2; 8.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
 DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 8.  
 DR Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;  
 KW DNA-binding; Repeat; Alternative splicing.  
 FT ZN\_FING 229 251 C2H2-TYPE 1.  
 FT ZN\_FING 257 279 C2H2-TYPE 2.  
 FT ZN\_FING 285 307 C2H2-TYPE 3.  
 FT ZN\_FING 318 340 C2H2-TYPE 4.  
 FT ZN\_FING 346 368 C2H2-TYPE 5.  
 FT ZN\_FING 374 398 C2H2-TYPE 6.  
 FT ZN\_FING 404 426 C2H2-TYPE 7.  
 FT ZN\_FING 434 456 C2H2-TYPE 8.  
 FT DOMAIN 462 524 GLN-RICH.  
 FT DOMAIN 467 506 ALA-RICH.  
 FT VARSPPLIC 103 118 Missing (in isoform 2).  
 FT VARSPPLIC 301 361 /FTId-VSP\_006921.  
 FT VARSPPLIC 361 379 /FTId-VSP\_006922.  
 FT CONFLICT 178 179 GG -> RS (IN REF. 1).  
 FT CONFLICT 576 577 LA -> WP (IN REF. 1).  
 SQ SEQUENCE 579 AA; 63139 MW; PBC242E0D1050C45 CRC64;  
 Query Match 60.0%; Score 48; DB 1; Length 579;  
 Best Local Similarity 90.0%; Pred. No. 13;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 5 LYREGGGGGG 14  
 Db 172 LTEGGGGGGG 181  
 RESULT 4  
 UVRA\_ECO57 STANDARD; PRT; 940 AA.  
 ID\_UVRA\_ECO57

AC Q8X5U9;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE UVRA system protein A (UvrA protein) (Excinuclease ABC subunit A).  
 GN UVRA OR Z5657 OR ECS5040.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.P., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RL Nature 409:529-533 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22 (2001).  
 CC -1- FUNCTION: The UvrABC repair system catalyzes the recognition and  
 CC processing of DNA lesions. UvrA is an ATPase and a DNA-binding  
 CC protein. A damage recognition complex composed of 2 uvrA and 2  
 CC uvrB subunits scans DNA for abnormalities. When the presence of a  
 CC lesion has been verified by uvrB, the uvrA molecules dissociate  
 CC (By similarity).  
 CC -1- SUBUNIT: Forms a heterotetramer with uvrB during the search for  
 CC lesions (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. UVRA SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AB005638; AAC59256.1; -  
 DR EMBL; AP002568; BAB38463.1; -  
 DR PIR; H91258; H91258.  
 DR HAMAP; MF 00205; -; 1.  
 DR InterPro; IPR003439; ABC transporter.  
 DR InterPro; IPR004602; UvrA.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR TIGRFAMS; TIGR00630; uvrA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 2.  
 DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW SOS response; Excision nuclease; DNA repair; DNA recombination;  
 KW DNA excision; ATP-binding; DNA-binding; Repeat; Zinc; Metal-binding;  
 KW Zinc-finger; Complete proteome.  
 FT NP\_BIND 31 38 ATP (POTENTIAL).  
 FT ZN\_FING 253 280 C4-TYPE.  
 FT NP\_BIND 640 647 ATP (POTENTIAL).  
 FT ZN\_FING 740 766 C4-TYPE.  
 SQ SEQUENCE 940 AA; 103884 MW; A20C90C935A0ACEB CRC64;  
 Query Match 60.0%; Score 48; DB 1; Length 940;

Best Local Similarity 64.3%; Pred. No. 21;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 LADLYEAGGGGGG 15  
DB 901 IVDLPGEGGGGG 914

RESULT 5  
UVR\_A ECOL6 STANDARD; PRT; 940 AA.  
AC Q8P802; P76788; PRT; 940 AA.  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE UVRABC system protein A (UvrA protein) (Excinuclease ABC subunit A).  
GN UVR\_A OR C5048  
OS Escherichia coli O6.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=217992;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=06:HI / CFT073 / ATCC 700928;  
RX MEDLINE=22388234; PubMed=124711157;  
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
RT "Extensive mosaic structure revealed by the complete genome sequence  
of uropathogenic Escherichia coli.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
CC -1- FUNCTION: The UVRABC repair system catalyzes the recognition and  
processing of DNA lesions. UvrA is an ATPase and a DNA-binding  
protein. A damage recognition complex composed of 2 uvrA and 2  
uvrB subunits scans DNA for abnormalities. When the presence of a  
lesion has been verified by uvrB, the uvrA molecules dissociate  
(By similarity).  
CC -1- SUBUNIT: Forms a heterotetramer with uvrB during the search for  
lesions. (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1- SUBFAMILY: BELONGS TO THE ABC TRANSPORTER FAMILY. UVR\_A SUBFAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF016770; AA083474.1; -  
DR HAMAP; MF\_00205; -; 1.  
DR InterPro; IPR003439; ABC transporter.  
DR InterPro; IPR004602; UvrA.  
DR Pfam; PF00005; ABC tran; 1.  
DR ProDom; PD000006; ABC transporter; 1.  
DR TIGRfam; TIGR00630; uvrA; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 2.  
DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
KW SOS response; Excision nuclease; DNA repair; DNA recombination;  
KW DNA excision; ATP-binding; DNA-binding; Repeat; Zinc; Metal-binding;  
KW Zinc-finger; Complete proteome.  
FT NP\_BIND 31 38 ATP (POTENTIAL).  
FT ZN\_FING 253 280 C4-TYPE.  
FT NP\_BIND 640 647 ATP (POTENTIAL).  
FT ZN\_FING 740 766 C4-TYPE.  
SQ SEQUENCE 940 AA; 103882 MW; A20C90C93816ACEB CRC64;

Query Match 60.0%; Score 48; DB 1; Length 940;  
Best Local Similarity 64.3%; Pred. No. 21;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 LADLYEAGGGGGG 15  
DB 901 IVDLPGEGGGGG 914

RESULT 6  
UVR\_A ECOLI STANDARD; PRT; 940 AA.  
AC P07671; P76788; PRT; 940 AA.  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE UvrABC system protein A (UvrA protein) (Excinuclease ABC subunit A).  
GN UVR\_A OR DINE OR B4058 OR SF4146.  
OS Escherichia coli, and  
OS Shigella flexneri.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562, 623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli;  
RX MEDLINE=8618204; PubMed=3007478;  
RA Husain I., van Houten B., Thomas D.C., Sancar A.;  
RT "Sequences of Escherichia coli uvrA gene and protein reveal two  
potential ATP binding sites.";  
RL J. Biol. Chem. 261:4895-4901(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli; STRAIN=K12 / MG1655;  
RX MEDLINE=9408932; PubMed=8265357;  
RA Blattner P.R., Burland V.D., Plunkett G. III, Sofia H.J.,  
RA Daniels D.L.;  
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the  
region from 89.2 to 92.8 minutes.";  
RL Nucleic Acids Res. 21:5408-5417(1993).  
RN [3]  
RP SEQUENCE OF 1-25 FROM N.A.  
RC SPECIES=E.coli;  
RX MEDLINE=8329251; PubMed=6310514;  
RA Backendorf C., Brandesa J.A., Kartasova T., van de Putte P.;  
RT "In vivo regulation of the uvrA gene: role of the '-10' and '-35'  
promoter regions.";  
RL Nucleic Acids Res. 11:5795-5810(1983).  
RN [4]  
RP SEQUENCE OF 1-14 FROM N.A.  
RC SPECIES=E.coli;  
RX MEDLINE=82220077; PubMed=6283374;  
RA Sancar A., Sancar G.B., Rupp W.D., Little J.W., Mount D.W.;  
RT "LexA protein inhibits transcription of the E. coli uvrA gene in  
vitro.";  
RL Nature 298:96-98(1982).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;  
RX MEDLINE=22272406; PubMed=12384590;  
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
RA Yu J.;  
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
through comparison with genomes of Escherichia coli K12 and O157.";  
RL Nucleic Acids Res. 30:4432-4441(2002).  
RN [6]  
RP CHARACTERIZATION.  
RC SPECIES=E.coli;  
RX MEDLINE=91208117; PubMed=1826851;  
RA Wyles G.M., Sancar A.;  
RT "Isolation and characterization of functional domains of UvrA.";  
RL Biochemistry 30:3834-3840(1991).  
RN [7]  
RP MUTAGENESIS OF CYS-253.

RC SPECIES=E.coli;  
 RX MEDLINE=89380205; PubMed=2550431;  
 RA Navaratnam S., Myles G.M., Strange R.W., Sancar A.;  
 RT "Evidence from extended X-ray absorption fine structure and site-specific mutagenesis for zinc fingers in UvrA protein of *Escherichia coli*.";  
 RT J. Biol. Chem. 264:16067-16071(1989).  
 CC -1- FUNCTION: The UvrABC repair system catalyzes the recognition and processing of DNA lesions. UvrA is an ATPase and a DNA-binding protein. A damage recognition complex composed of 2 uvrA and 2 uvrB subunits scans DNA for abnormalities. When the presence of a lesion has been verified by uvrB, the uvrA molecules dissociate (By similarity).  
 CC -1- SUBUNIT: Forms a heterotetramer with uvrB during the search for lesions (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- MISCELLANEOUS: Binds about 2 zinc atoms/molecule.  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. UVR A SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; M13495; AAA24754.1; -  
 CC EMBL; U00006; AAC3152.1; -  
 CC EMBL; AE000479; AAC77028.1; -  
 CC EMBL; X01621; CAA25764.1; -  
 CC EMBL; J01721; AAA24753.1; -  
 CC EMBL; AE015423; AAA45568.1; -  
 CC PIR; A23865; BVBCUA.  
 CC ECODBASE; H124.0; 6TH EDITION.  
 CC Scogene; EG11061; uvrA.  
 CC HAMAP; MF 00205; -; 1.  
 CC InterPro; IPR003439; ABC transporter.  
 CC InterPro; IPR004602; UvrA.  
 CC Pfam; PF00005; ABC\_tran; 1.  
 CC TIGRfam; TIGR00606; ABC transporter; 1.  
 CC PROSITE; PS00211; ABC\_TRANSPORTER\_1; 2.  
 CC PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW SOS response; Excision nuclease; DNA repair; DNA recombination; DNA excision; ATP-binding; DNA-binding; Repeat; Zinc; Metal-binding; Zinc-finger; Complete proteome.  
 KW NP\_BIND 31 38 ATP.  
 FT ZN\_BIND 253 280 C4-TYPE.  
 FT NP\_BIND 640 647 ATP.  
 FT ZN\_BIND 740 766 C4-TYPE.  
 FT MUTAGEN 253 253 C->A,H,S. REDUCED ACTIVITY.  
 FT SEQUENCE 940 AA; 103867 MW; D61AAAE6514B860C CRC64;  
 Qy Query Match 60.0%; Score 48; DB 1; Length 940;  
 Db Best Local Similarity 64.3%; Pred. No. 21;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 Qy 2 LADLYEGGGGGG 15  
 Db 901 IVDLGGGGGGGG 914  
 RESULT 7  
 ID UVR A SALTY STANDARD; PRT; 941 AA.  
 AC P37434.  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE UvrABC system protein A (UvrA protein) (Excinuclease ABC subunit A).  
 GN UVR A OR STM4254 OR STY4450 OR T4160.  
 OS Salmonella typhimurium, and

OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602, 601;  
 RN [1] RN  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhimurium; STRAIN=NM522;  
 RA Alberti M., Li Y.F., Sancar A., Hearst J.E.;  
 RL Submitted (AUG-1992) to the EMBL/GenBank/DBJ databases.  
 RN [2] RN  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;  
 RA "Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2.";  
 RT Nature 413:852-856(2001).  
 RL [3] RL  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhi; STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M., Baker S., Basham D., Brooks K., Chillingworth T., Connor P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;  
 RA "Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18.";  
 RT Nature 413:848-852(2001).  
 RL [4] RL  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhi; STRAIN=TY2 / ATCC 700931;  
 RX MEDLINE=22531367; PubMed=12644504;  
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.P., Rose D.J., Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;  
 RA "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18.";  
 RT J. Bacteriol. 185:2330-2337(2003).  
 CC -1- FUNCTION: The UvrABC repair system catalyzes the recognition and processing of DNA lesions. UvrA is an ATPase and a DNA-binding protein. A damage recognition complex composed of 2 uvrA and 2 uvrB subunits scans DNA for abnormalities. When the presence of a lesion has been verified by uvrB, the uvrA molecules dissociate (By similarity).  
 CC -1- SUBUNIT: Forms a heterotetramer with uvrB during the search for lesions (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- MISCELLANEOUS: Binds about 2 zinc atoms/molecule (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. UVR A SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; M93014; AAA27250.1; -  
 CC EMBL; AE008898; AAL23078.1; -  
 CC EMBL; AL627282; CAD09238.1; -  
 CC EMBL; AE016848; AA071624.1; -  
 CC StyGene; SG10413; uvrA.  
 CC HAMAP; MF 00205; -; 1.  
 CC InterPro; IPR003439; ABC transporter.  
 CC InterPro; IPR004602; UvrA.



DR Pfam; PF00005; ABC tran; 2.  
 DR ProDom; PD000006; ABC transporter; 1.  
 DR TIGRFAMs; TIGR00630; uvra; 1.  
 DR PROSITE; PS00211; ABC TRANSPORTER 1; 2.  
 DR PROSITE; PS00893; ABC TRANSPORTER 2; 1.  
 KW SOS response; Excision nuclease; DNA repair; DNA recombination;  
 KW DNA excision; ATP-binding; DNA-binding; Repeat; Zinc; Metal-binding;  
 KW Zinc-finger; Complete proteome.  
 FT NP\_BIND 31 38 ATP  
 FT ZN\_FING 253 280 C4-TYPE  
 FT NP\_BIND 640 647 ATP  
 FT ZN\_FING 740 766 C4-TYPE  
 SQ SEQUENCE 941 AA; 103928 MW; C4AFC9F549060C26 CRC64;  
 Query Match 60.0%; Score 48; DB 1; Length 941;  
 Best Local Similarity 64.3%; Pred. No. 21;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 LADLYEEGGGGGGE 15  
 Db 901 IVDLPGEGGGGGGE 914  
 RESULT 8  
 ID UVRA\_YERPE STANDARD; PRT; 947 AA.  
 AC Q8ZJ07;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DE UvrABC system protein A (UvrA protein) (Excinuclease ABC subunit A).  
 GN UVRA OR YP00324 OR Y0580.  
 OS *Yersinia pestis*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; *Yersinia*.  
 OX NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 / Biovar Orientalis;  
 RX MEDLINE=21470413; PubMed=11586360;  
 RA Parthill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
 RA Prentice M.B., Sebahia M., James K.D., Church C., Mungall K.L.,  
 RA Baker S., Baeham D., Bentley S., Davies R.M., Brooks K., Cardeno-Tarraga A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Fellwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;  
 RT "Genome sequence of *Yersinia pestis*, the causative agent of plague."  
 RL Nature 413:523-527(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KIM5 / Biovar Mediaevalis;  
 RX MEDLINE=22137863; PubMed=12142430;  
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,  
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,  
 RA Straley S.C., McDonough K.A., Milles M.L., Matson J.S., Blattner F.R.,  
 RA Perry R.D.;  
 RT "Genome sequence of *Yersinia pestis* KIM5."  
 RL J. Bacteriol. 184:4601-4611(2002).  
 CC -1- FUNCTION: The UvrABC repair system catalyzes the recognition and  
 processing of DNA lesions. UvrA is an ATPase and a DNA-binding  
 protein. A damage recognition complex composed of 2 uvrA and 2  
 uvrB subunits scans DNA for abnormalities. When the presence of a  
 lesion has been verified by uvrB, the uvrA molecules dissociate  
 (By similarity).  
 CC -1- SUBUNIT: Forms a heterotetramer with uvrB during the search for  
 lesions (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. UVRA SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC  
 DR EMBL; AJ414142; CAC89185.1; --  
 DR EMBL; AB013660; AAM84168.1; --  
 DR HAMAP; MF\_00205; --; 1.  
 DR InterPro; IPR003439; ABC transporter.  
 DR InterPro; IPR004602; UvrA.  
 DR Pfam; PF00005; ABC tran; 1.  
 DR ProDom; PD000006; ABC transporter; 1.  
 DR TIGRFAMs; TIGR00630; uvra; 1.  
 DR PROSITE; PS00211; ABC TRANSPORTER 1; 2.  
 DR PROSITE; PS00893; ABC TRANSPORTER 2; 1.  
 KW SOS response; Excision nuclease; DNA repair; DNA recombination;  
 KW DNA excision; ATP-binding; DNA-binding; Repeat; Zinc; Metal-binding;  
 KW Zinc-finger; Complete proteome.  
 FT NP\_BIND 31 38 ATP (POTENTIAL).  
 FT ZN\_FING 253 280 C4-TYPE.  
 FT NP\_BIND 640 647 ATP (POTENTIAL).  
 FT ZN\_FING 740 766 C4-TYPE.  
 SQ SEQUENCE 947 AA; 104241 MW; 44B38B4F133410FD CRC64;  
 Query Match 60.0%; Score 48; DB 1; Length 947;  
 Best Local Similarity 64.3%; Pred. No. 21;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 LADLYEEGGGGGGE 15  
 Db 901 IVDLPGEGGGGGGE 914  
 RESULT 9  
 ID YR21\_TRSVR STANDARD; PRT; 201 AA.  
 AC P25245;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical 20.2 kDa protein in RNA2.  
 OS Tomato ringspot virus (isolate raspberry) (TOMRSV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;  
 OC Nepovirus.  
 OX NCBI\_TaxID=12281;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=91311402; PubMed=1856689;  
 RA Rott M.B., Tremaine J.H., Rochon D.M.;  
 RT "Nucleotide sequence of tomato ringspot virus RNA-2."  
 RL J. Gen. Virol. 72:1505-1514(1991).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC  
 DR EMBL; D12477; BAA02044.1; --  
 DR PIR; JQ1094; JQ1094.  
 KW Hypothetical protein.  
 FT DOMAIN 15 22 POLY-GLY.  
 FT DOMAIN 61 66 POLY-GLY.  
 FT DOMAIN 144 148 POLY-GLY.  
 SQ SEQUENCE 201 AA; 20194 MW; 9038506E18D7B450 CRC64;  
 Query Match 58.8%; Score 47; DB 1; Length 201;  
 Best Local Similarity 75.0%; Pred. No. 6;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;



DR InterPro; IPR000595; CNMP binding.  
 DR InterPro; IPR005821; Ion trans.  
 DR InterPro; IPR005820; M-channel nlg.  
 DR Pfam; PF00027; CNMP binding; 1.  
 DR Pfam; PF00520; ion trans; 1.  
 DR SMART; SM0100; CNMP; 1.  
 DR PROSITE; PS00888; CNMP BINDING 1; 1.  
 DR PROSITE; PS00889; CNMP BINDING 2; FALSE\_NEG.  
 DR PROSITE; PS00442; CNMP BINDING 3; 1.  
 DR Transmembrane; Transmembrane; Glycophorin;  
 KW Potassium channel; Potassium; Potassium transport; Sodium transport;  
 KW CAMP; CAMP-binding; Transmembrane; Glycophorin;  
 FT DOMAIN 1 267  
 FT TRANSMEM 268 288  
 FT TRANSMEM 295 315  
 FT DOMAIN 316 341  
 FT TRANSMEM 342 362  
 FT TRANSMEM 370 390  
 FT TRANSMEM 391 421  
 FT TRANSMEM 422 442  
 FT TRANSMEM 466 487  
 FT TRANSMEM 498 518  
 FT DOMAIN 519 1175  
 FT DOMAIN 210 261  
 FT  
 FT DOMAIN 68 127  
 FT NP\_BIND 596 711  
 FT CAMP.  
 FT DOMAIN 796 1160  
 FT CARBOHYD 459 459  
 FT SEQUENCE 1175 AA; 126141 MW; 35A75FA9C710BD69 CRC64;  
 SQ  
 Query Match 57.5%; Score 46; DB 1; Length 1175;  
 Best Local Similarity 64.3%; Pred. No. 51;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 ALADYERGGGGG 14  
 DB 105 SLASLGRGGGGG 118  
 RESULT 12  
 ID BMP7 HUMAN  
 AC P18075; Q9H512; Q9NT07; PRT; 431 AA.  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Bone morphogenetic protein 7 precursor (BMP-7) (Osteogenic protein 1) (OP-1).  
 EN BMP7 OR OPI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Placenta;  
 RX MEDLINE=90291971; PubMed=2357959;  
 RA Ozkaynak B., Rueger D.C., Drier B.A., Corbett C., Ridge R.J.,  
 RA Sampath T.K., Oppermann H.;  
 RT "OP-1 cDNA encodes an osteogenic protein in the TGF-beta family.";  
 RL EMBO J. 9:2085-2093(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9108608; PubMed=2263636;  
 RA Celeste A.J., Iannazi J.A., Taylor R.C., Hewick R.M., Rosen V.,  
 RA Wang E.A., Wozney J.M.;  
 RT "Identification of transforming growth factor beta family members  
 RT present in bone-inductive protein purified from bovine bone.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9843-9847(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21638749; PubMed=11780052;

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Clee C.M.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Leharvaish M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConnachie L.J., McEay K., McMurray A.A.,  
 RA Milne S.A., Mistry D., Moore M.J.P., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore A.C., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Heish F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting J., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green B.D., Dickinson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 293-431.  
 RX MEDLINE=96149402; PubMed=8570652;  
 RA Griffith D.L., Keck P.C., Sampath T.K., Rueger D.C., Carlson W.D.;  
 RT "Three-dimensional structure of recombinant human osteogenic protein  
 RT 1: structural paradigm for the transforming growth factor beta  
 RT superfamily";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:878-883(1996).  
 CC -1- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION. MAY BE THE  
 CC OSTEOINDUCTIVE FACTOR RESPONSIBLE FOR THE PHENOMENON OF  
 CC EPITHELIAL OSTEOGENESIS. PLAYS A ROLE IN CALCIUM REGULATION  
 CC AND BONE HOMEOSTASIS.  
 CC -1- SUBUNIT: Homodimer; disulfide-linked.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE KIDNEYS AND BLADDER. LOWER  
 CC LEVELS SEEN IN THE BRAIN.  
 CC -1- SIMILARITY: Belongs to the TGF-beta family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC EMBL; X51801; CAA36100.1; --
CC EMBL; M60316; AAA36738.1; --
CC EMBL; AL122058; CAB90273.1; --
CC EMBL; AL157414; CAC08434.1; --
CC EMBL; BC008584; AAO08584.1; --
CC PIR; C39263; BMH07.
CC PDB; 1EMP; 23-JUL-97.
CC PDB; 1MAU; 18-DEC-02.
CC Genew; HGNC:1074; BMP7.
CC MIM; 112267; --
CC GO; GO:0001501; P:skeletal development; TAS.
CC InterPro; IPR001839; TGFb.
CC InterPro; IPR001111; TGFb.N.
CC Pfam; PF00019; TGFb-beta; 1.
CC Pfam; PF00688; TGFb-propeptide; 1.
CC ProDom; PD000357; TGFb; 1.
CC SMART; SM00204; TGFb; 1.
CC PROSITE; PS00250; TGF_BETA_1; 1.
CC Signal; Growth factor; Cytokine; Bone; Cartilage; Glycoprotein;
KW 3D-structure.
CC SIGNAL 1 29
CC PROPEP 30 292
CC CHAIN 293 431
CC DISULFID 330 396
CC DISULFID 339 428
CC DISULFID 363 430
CC DISULFID 395 395
CC CARBOHYD 187 187
CC CARBOHYD 302 302
CC CARBOHYD 321 321
CC CARBOHYD 372 372
CC STRAND 329 329
CC STRAND 331 333
CC STRAND 336 338
CC HELIX 339 342
CC TURN 343 347
CC STRAND 348 350
CC STRAND 353 355
CC STRAND 358 360
CC STRAND 362 362
CC TURN 370 371
CC HELIX 375 386
CC STRAND 388 390
CC STRAND 396 409
CC HELIX 411 413
CC STRAND 415 431
CC SEQUENCE 431 AA; 49313 MW; 47A05B45C6815F8A CRC64;

Query Match 56.9%; Score 45.5; DB 1; Length 431;
Best Local Similarity 52.6%; Pred No. 21;
Matches 10; Conservative 2; Mismatches 2; Indels 5; Gaps 1;

QY 2 LADLY-----EKGGGGGG 15
DB 83 MLDLYNAMAVEGGGPGGQ 101

RESULT 13
R887 DROME
ID RB87 DROME STANDARD; PRT; 386 AA.
AC P48810;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Heterogeneous nuclear ribonucleoprotein 87F (HRP36.1 protein) (P11
DE protein).
DE HRB87F OR HRP36.
DS Drosophila melanogaster (Fruit fly).
DC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
DC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Spheroidea; Drosophilidae; Drosophila.
OX NCBI TaxID=7227;
RN [1] SEQUENCE FROM N.A.
RP STRAIN=Oregon-R, and Canton-S; TISSUE=Ovary;
RX MEDLINE=9187645; PubMed=1849257;
RA Haynes S.R., Johnson D., Raychaudhuri G., Beyer A.L.;
RT "The Drosophila Hrb87F gene encodes a new member of the A and B hnRNP
RT protein group.";
RL Nucleic Acids Res. 19:25-31(1991).
RN [2] SEQUENCE FROM N.A.
RP STRAIN=Canton-S;
RX MEDLINE=92112968; PubMed=1730754;
RA Matunis E.L., Matunis M.J., Dreyfuss G.;
RT "Characterization of the major hnRNP proteins from Drosophila
RT melanogaster";
RL J. Cell Biol. 116:257-269(1992).
RN [3] SEQUENCE FROM N.A.
RP STRAIN=Canton-S; TISSUE=Embryo;
RX MEDLINE=92020124; PubMed=1717937;
RA Hovemann B.T., Dessen E., Mechler H., Mack E.;
RT "Drosophila snRNP associated protein P11 which specifically binds to
RT heat shock puff 93D reveals strong homology with hnRNP core protein
RT A1.";
RL Nucleic Acids Res. 19:4909-4914(1991).
CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF RIBONUCLEOSOMES. COULD BE
CC NEEDED TO ORGANIZE A CONCENTRATION GRADIENT OF A DORSALIZING
CC MORPHOGEN (DM) ORIGINATING IN THE GERMAL VESICLE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR AND/OR CYTOPLASMIC.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=P48810-1; Sequence=Displayed;
CC Name=2; Synonyms=HRP36.1;
CC IsoId=P48810-2; Sequence=VSP_005807;
CC -1- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

Query Match

56.2%; Score 45; DB 1; Length 386;

```

Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 0; Gaps 0;
QY 6 YBEGGGGGG 14
DB 343 YSQGGGGGG 351

RESULT 14
YJEB YEAST
ID YJEB YEAST STANDARD; PRT; 396 AA.
AC P47049.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein YJL048C.
GN YJL048C OR J1164.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
ON NCBI_TaxID=4932;
RX SEQUENCE FROM N.A. (ISOFORM 1).
RA Pohl T.M., Aljinovic G.;
RA Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: Contains 1 UBX domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; Z49323; CA899339.1;
DR F1R; S56820; S56820.
DR SGB; S0003584; YJL048C.
DR InterPro; IPR001012; UBX.
DR Pfam; PF00789; UBX; 1.
DR SMART; SM00166; UBX; 1.
DR PROSITE; PS50033; UBX; 1.
DR Hypothetical protein.
FT DOMAIN 186 264
FT SEQUENCE 396 AA; 45016 MW; 42C6217D2B52B425 CRC64;
BX.
Query Match 56.2%; Score 45; DB 1; Length 396;
Best Local Similarity 63.6%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
DY 4 DLYEGGGGGG 14
DB 331 DWSDGGGGGG 341

RESULT 15
JAZ DROME
ID JAZ DROME STANDARD; PRT; 404 AA.
AC Q27294; Q24445; Q9VXI4.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RNA-binding protein cabeza (Sarcopa-associated RNA-binding fly
DE homolog) (P19).
OS CAZ OR SARPH OR CG3606.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RX SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RA STRAIN=Canton-S;

```

```

MEDLINE=95349623; PubMed=7623847;
RA Immanuel D., Zinszner H., Ron D.;
RA "Association of SARPH (sarcoma-associated RNA-binding fly homolog)
RA with regions of chromatin transcribed by RNA polymerase II.";
RL Mol. Cell. Biol. 15:4562-4571 (1995).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=Canton-S;
RX MEDLINE=95223793; PubMed=7708500;
RA Stolow D.T., Haynes S.R.;
RA "Cabeza, a Drosophila gene encoding a novel RNA binding protein,
RA shares homology with EWS and TLS, two genes involved in human sarcoma
RA formation.";
RL Nucleic Acids Res. 23:835-843 (1995).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin N.L., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [4]
RP SEQUENCE OF 39-404 FROM N.A.
RC STRAIN=Oregon-R;
RA Haynes S.R.;
RA Submitted (APR-1988) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 212-261 FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=87175568; PubMed=3031652;
RA Haynes S.R., Rabbert M.L., Mozer B.A., Forquignon F., Dawid I.B.;
RA "Pen repeat sequences are GGN clusters and encode a glycine-rich
RA domain in a Drosophila cDNA homologous to the rat helix destabilizing
RA protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:1819-1823 (1987).
CC -1- FUNCTION: MAY PARTICIPATE IN A FUNCTION COMMON TO THE EXPRESSION
CC OF MOST GENES TRANSCRIBED BY RNA POLYMERASE II.
CC -1- SUBCELLULAR LOCATION: Nuclear.

```

```

CC CC -1- ALTERNATIVE PRODUCTS:
CC CC Event=Alternative splicing; Named isoforms=2;
CC CC Comment=Additional isoforms seem to exist;
CC CC Name=2;
CC CC IsoId=Q27294-1; Sequence=Displayed;
CC CC Name=1;
CC CC IsoId=Q27294-2; Sequence=VSP_005778;
CC CC Note=No experimental confirmation available;
CC CC -1- TISSUE SPECIFICITY: UBIQUITOUS. ENRICHED IN THE BRAIN AND CENTRAL
CC CC NERVOUS SYSTEM DURING EMBRYOGENESIS. ENRICHED IN THE ADULT HEAD.
CC CC EMBRYOS CONTAIN BOTH TYPE 1 AND TYPE 2 ISOFORMS, WHEREAS LATER IN
CC CC DEVELOPMENT (HEADS AND TORSOS) ONLY THE TYPE 2 ISOFORM IS
CC CC DETECTED.
CC CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE DEVELOPING EMBRYO FROM THE
CC CC EARLIEST STAGES OF CELLULARIZATION AND IS SUBSEQUENTLY FOUND IN
CC CC MANY CELL TYPES.
CC CC -1- MISCELLANEOUS: 'CABEZA' MEANS 'HEAD' IN SPANISH.
CC CC -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC CC -1- SIMILARITY: Contains 1 RANBP2-type zinc finger.
CC CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; U13178; AA86955.1; -
CC CC EMBL; L37083; AAC41563.1; -
CC CC EMBL; AE003501; AAF48578.1; -
CC CC EMBL; M15765; AAT70425.1; -
CC CC F01; S54729; S54729.
CC CC FlyBase; FBgn0011571; caz.
CC CC InterPro; IPR000504; RNA rec mot.
CC CC InterPro; IPR001876; Znf_RanGDP.
CC CC Pfam; PF00076; rtm; 1.
CC CC Pfam; PF00641; zf-RanBP; 1.
CC CC SMART; SM00360; RRM; 1.
CC CC SMART; SM00547; Znf_RBZ; 1.
CC CC PROSITE; PS0102; RRM; 1.
CC CC PROSITE; PS00030; RRM RNP 1; FALSE_NEG.
CC CC PROSITE; PS01358; ZF_RANBP2_1; 1.
CC CC PROSITE; PS0199; ZF_RANBP2_2; 1.
CC CC Nuclear protein; Zinc-finger; Metal-binding; RNA-binding;
CC CC Alternative splicing.
CC CC DOMAIN 42 111 GLY-RICH.
CC CC FT DOMAIN 119 205 RNA-BINDING (RRM).
CC CC FT DOMAIN 212 275 GLY-RICH.
CC CC FT ZN_FING 280 309 RANBP2-TYPE.
CC CC FT DOMAIN 312 391 GLY-RICH.
CC CC FT VARSPLIC 4 47 Missing (in isoform 1).
CC CC FT CONFLICT 39 41 PNY -> LFI (IN REF. 4).
CC CC FT CONFLICT 92 92 P -> H (IN REF. 3).
CC CC FT CONFLICT 108 108 G -> GG (IN REF. 3).
CC CC FT CONFLICT 253 258 MISSING (IN REF. 3).
CC CC FT CONFLICT 283 283 D -> E (IN REF. 4 AND 5).
CC CC FT CONFLICT 389 398 DGGPMRDGG -> MVDOEKWS (IN REF. 4).
CC CC SEQUENCE 404 AA; 39141 MW; 7062A0446BEA5984 CRC64;

```

Query Match 56.2%; Score 45; DB 1; Length 404;

Best Local Similarity 77.8%; Pred.No. 23;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 6 YEEGGGGG 14

Db 262 YDRGGGGG 270

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2003, 11:56:53 ; Search time 35 Seconds  
(without alignments)  
110.594 Million cell updates/sec

Title: US-09-807-610-11

Perfect score: 80

Sequence: % 1 ALADYEEGGGGGGE 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organalle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	85.0	20	Q9UP87	Q9up87 homo sapien
2	51	63.7	213	Q96IP9	Q96ip9 homo sapien
3	51	63.7	258	Q9UKY7	Q9uky7 homo sapien
4	51	63.7	316	Q994A5	Q994a5 unidentified
5	51	63.7	396	Q8GSV5	Q8gsy5 oryza sativ
6	51	63.7	422	Q8S1G8	Q8slg8 oryza sativ
7	51	63.7	427	Q8GAW7	Q8gxw7 arabidopsis
8	51	63.7	1357	Q8Q334	Q8q334 chimpanzee
9	50	62.5	294	Q8RVV9	Q8rvv9 rhodospiril
10	49.5	61.9	362	Q9L7U6	Q9l7u6 bos taurus
11	49	61.3	206	O44886	O44886 caenorhabdi
12	49	61.3	591	Q95L39	Q95l39 bos taurus
13	48	60.0	67	Q8H3X2	Q8h3x2 oryza sativ
14	48	60.0	141	Q9PVJ4	Q9pvj4 varanus dum
15	48	60.0	148	Q8VXY1	Q8vxy1 arabidopsis
16	48	60.0	179	Q8H2G0	Q8h2g0 oryza sativ

17	48	60.0	256	16	Q92SS3	Q92s83 rhizobium m
18	48	60.0	325	10	Q9C9L2	Q9c9l2 arabidopsis
19	48	60.0	397	11	Q9EQJ1	Q9eqj1 rattus norv
20	48	60.0	517	11	Q8C3E0	Q8c3e0 mus musculu
21	48	60.0	631	10	Q8LIP4	Q8lip4 oryza sativ
22	48	60.0	940	16	Q8FB02	Q8fb02 escherichia
23	48	60.0	947	16	Q8ZJ07	Q8zj07 yersinia pe
24	48	60.0	950	15	Q8EA79	Q8ear79 shewanella
25	47	58.8	62	13	Q8JHZ3	Q8jhz3 ostracion s
26	47	58.8	127	11	P70370	P70370 mus musculu
27	47	58.8	362	10	Q8LR17	Q8lr17 oryza sativ
28	47	58.8	576	17	Q8TM20	Q8tm20 methanosarc
29	47	58.8	678	5	Q9VBW6	Q9vbw6 drosophila
30	47	58.8	726	16	Q9IOF3	Q9iof3 pseudomonas
31	47	58.8	1094	10	Q9LLI2	Q9lll2 zea mays (m
32	47	58.8	4655	5	Q9VH97	Q9vh97 drosophila
33	46.5	58.1	158	10	Q9C618	Q9c618 arabidopsis
34	46.5	58.1	158	10	Q8L9C8	Q8l9c8 arabidopsis
35	46.5	58.1	455	10	Q9C752	Q9c752 arabidopsis
36	46	57.5	40	2	Q9X968	Q9x968 salmonella
37	46	57.5	107	10	Q8S1H2	Q8slh2 oryza sativ
38	46	57.5	137	10	Q8L536	Q8l536 oryza sativ
39	46	57.5	156	11	Q8CBN4	Q8cbn4 mus musculu
40	46	57.5	184	10	Q8S202	Q8s202 oryza sativ
41	46	57.5	257	16	Q9XK57	Q9xk57 streptococc
42	46	57.5	257	16	Q8NZ62	Q8nz62 streptococc
43	46	57.5	257	16	Q8KGN3	Q8kgn3 streptococc
44	46	57.5	258	16	Q8E745	Q8e745 streptococc
45	46	57.5	258	16	Q8E1M9	Q8elm9 streptococc

#### ALIGNMENTS

RESULT 1  
Q9UP87 PRELIMINARY, PRT; 20 AA.  
AC Q9UP87; 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
DE Type II interleukin-1 receptor antagonist (Fragment).  
GN IL-1RA3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Muzio M., Polencarutti N., Sironi M., Transidico P., Introna M.,  
RA Mantovani A.;  
RT "Characterization of the type II intracellular IL-1 receptor  
RT antagonist (IL-1ra3): a depot IL-1ra without demonstrable  
RT intracellular function."  
RU Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF057168; AAC13499.1;  
KW Receptor.  
FT NON TER 1 1  
SQ NON TER 20 20  
SQ SEQUENCE 20 AA; 1956 MW; 47DDD28AE94667D6 CRC64;  
Query Match 85.0%; Score 68; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred.No.0.001;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 4 DLVEEGGGGGGE 15  
|||  
Db 1 DLVEEGGGGGGE 12  
RESULT 2  
Q96IP9 PRELIMINARY, PRT; 213 AA.  
ID Q96IP9

```

Db      251  IAQVYGGGGGGG 263
      :|:|:|:|:|
RESULT 10
Q9L7U6
ID      Q9L7U6      PRELIMINARY;      PRT;      362 AA.
AC      Q9L7U6;
DT      01-OCT-2000 (TREMBLrel. 15, Created)
DT      01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      Polar motor switch protein PflM.
GN      PflM.
OS      Rhodospirillum centenum (Rhodocista centenaria)
OC      Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC      Rhodospirillaceae; Rhodospirillum.
OX      NCBI_TaxID=34018;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ATCC51521;
RA      McClain J.R., Rollo D., Bauer C.E., Rushing B.R., Jiang Z.;
RT      "Gene duplication and lateral transfer events giving rise to
RT      Rhodospirillum centenum polar and lateral flagellar motor switch
RT      components."
RL      Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF220001; AAF35843.1;
DR      InterPro; IPR001589; Pfam_PflM.
DR      InterPro; IPR001543; Spoa.
DR      Pfam; PF02154; PflM; 1.
DR      Pfam; PF01052; Spoa; 1.
DR      TIGRFAMs; TIGR01397; PflM switch; 1.
SQ      SEQUENCE 362 AA; 39975 MW; 7C81B6D1121AB7A6 CRC64;

Query Match      61.9%; Score 49.5; DB 2; Length 362;
Best Local Similarity 73.3%; Pred. No. 16;
Matches 11; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

QY      * 1 ALADLYEGGGGGG 15
      ||| ||| ||| |||
Db      18 ALAD---EGGGGGG 29.

RESULT 11
O44886
ID      O44886      PRELIMINARY;      PRT;      206 AA.
AC      O44886;
DT      01-JUN-1998 (TREMBLrel. 06, Created)
DT      01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      Y8G1A.1 protein.
GN      Y8G1A.1.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OC      Rhabditidae; Pelodierinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Bristol N2;
RA      MEDLINE=94150718; PubMed=7906398;
RA      Wilson R., Alnsough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA      Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA      Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA      Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA      Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA      Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA      Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA      Smaildon N., Smith A., Sonhammer E., Staden R., Sulston R.,
RA      Thierry-Mieg J., Thomas K., Vaughan M., Vaughan K., Waterston R.,
RA      Watson A., Weinstock L., Wilkinson-Sproat J., Wooldman P.;
RT      "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT      elegans."
UL      Nature 368:32-38(1994).
RN      [2]

```

```

RP      SEQUENCE FROM N.A.
RC      STRAIN=Bristol N2;
RA      Cordes M., Le T.T.;
RT      "The sequence of C. elegans cosmid Y8G1A."
RL      Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Bristol N2;
RA      Waterston R.;
RL      Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF040656; AAB35048.1;
DR      WormPep; Y8G1A.1; CE18355.
DR      InterPro; IPR004296; DUF236.
DR      Pfam; PF03057; DUF236; 2.
SQ      SEQUENCE 206 AA; 20378 MW; 09A8477DB7962F4E CRC64;

Query Match      61.3%; Score 49; DB 5; Length 206;
Best Local Similarity 63.6%; Pred. No. 10;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      4 DLYEEGGGGGGG 14
      ||| ||| ||| |||
Db      156 DIPKGGGGGGG 166

RESULT 12
Q95L39
ID      Q95L39      PRELIMINARY;      PRT;      591 AA.
AC      Q95L39;
DT      01-DEC-2001 (TREMBLrel. 19, Created)
DT      01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      Lysyl oxidase-like 1.
GN      LOXL1.
OS      Bos taurus (Bovine).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC      Bovidae; Bovinae; Bos.
OX      NCBI_TaxID=9913;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Skin;
RA      Borel A., Eischenberger D., Farjanel J., Kessler E., Gleyzal C.,
RA      Hulmes D., Sommer P., Pont B.;
RT      "Lysyl oxidase-like 1 from bovine aorta."
RL      Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF421185; AAL1312.1;
DR      InterPro; IPR001695; Lysyl oxidase.
DR      Pfam; PF01186; Lysyl oxidase; 1.
DR      PRINTS; PR00074; LYSYLOXIDASE.
DR      ProDom; PD013887; Lysyl oxidase; 1.
DR      PROSITE; PS00926; LYSYL OXIDASE; 1.
SQ      SEQUENCE 591 AA; 64496 MW; 337A344E64F0FC73 CRC64;

Query Match      61.3%; Score 49; DB 6; Length 591;
Best Local Similarity 90.0%; Pred. No. 33;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      6 YEEGGGGGGG 15
      ||| ||| ||| |||
Db      231 YEEYGGGGGGG 240

RESULT 13
Q8H3X2
ID      Q8H3X2      PRELIMINARY;      PRT;      67 AA.
AC      Q8H3X2;
DT      01-MAR-2003 (TREMBLrel. 23, Created)
DT      01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      P0585H11.5 protein.
GN      P0585H11.5.
OS      Oryza sativa (japonica cultivar-group).

```



```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Shariatideae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC
RL clone:PO585H11.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004342; BAC20738.1; -
SQ SEQUENCE 67 AA; 6790 MW; C356B491ECCE7E67 CRC64;

Query Match 60.0%; Score 48; DB 10; Length 67;
Best Local Similarity 69.2%; Pred. No. 4.2;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 LADYEEGGGGG 14
DB 49 LPDLRHGGGGG 61

RESULT 14
Q9PVJ4 PRELIMINARY; PRT; 141 AA.
AC Q9PVJ4
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Transcription factor HOXA13 (Fragment)
GN HOXA13.
OS Varanus dumerilli.
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Anguilliformia; Varanidae; Varanus.
OX NCBI_TaxID=62039;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20122168; PubMed=10656931;
RA Mortlock D.P., Sateesh P., Innis J.W.;
RT "Evolution of N-terminal sequences of the vertebrate HOXA13 protein.";
RL Mamm. Genome 11:151-158(2000).
DR EMBL; AF083102; AAD54647.1; -
FT NON_TER 1
FT NON_TER 141
SQ SEQUENCE 141 AA; 13913 MW; 840710CB2119BB98 CRC64;

Query Match 60.0%; Score 48; DB 13; Length 141;
Best Local Similarity 57.9%; Pred. No. 9.6;
Matches 11; Conservative 1; Mismatches 3; Indels 4; Gaps 1;

QY 1 ALADYEE-----GGGGGG 15
DB 118 AFADKYMTDTSVGGGGGG 136

RESULT 15
Q8VXY1 PRELIMINARY; PRT; 148 AA.
AC Q8VXY1
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein.
GN AT5G28630.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosida II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,

```

```

RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shimizu P., Southwick A., Shinozaki K.,
RA Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene AT5G28630 (GI:15241873).";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volkovskiy N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RL annotation.";
RN Genome Biol. 0:0-0(2002).
RP [3]
RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY074364; AAL67060.1; -
DR EMBL; AY086339; AAM64407.1; -
KW Hypothetical protein.
SQ SEQUENCE 148 AA; 16023 MW; 0DFPFE1E84409B00 CRC64;

Query Match 60.0%; Score 48; DB 10; Length 148;
Best Local Similarity 88.9%; Pred. No. 10;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 7 EREGGGGGG 15
DB 119 EREGGGGGG 127

Search completed: December 30, 2003, 12:02:47
Job time : 37 secs

```